# *RADIO TECHNOLOGY*

**TELEGRAPHY TELEPHONY TELEVISION**

**TRANSCRIPTION FACSIMILE**

*by*


ERNEST J. VOGT

*Consulting Radio Engineer*

*Assistant Chief Engineer, Aerojet Engineering Corporation*



PITMAN PUBLISHING CORPORATION

NEW YORK LONDON

PUBLISHED 1949



1.1

*Associated Companies*
SIR ISAAC PITMAN & SONS, LTD.
London Melbourne Johannesburg
SIR ISAAC PITMAN & SONS (CANADA), LTD.
Toronto

***To Margaret Haley Vogt***

# PREFACE

The study of radio is essentially a study of the peculiar behavior of alternating currents at extremely high frequencies. For a comprehensive understanding of these high-frequency phenomena, it is prerequisite that the student be thoroughly cognizant of the behavior of alternating currents at low or commercial frequencies.

In this book, the author has included the necessary direct- and alternating-current theory to prepare the student for a proper concept of the basic radio phenomena that are presented as the text progresses. The earlier chapters present nomenclature, axiomatic definitions, and the essential mathematics of radio. The reader is therefore armed at the outset with the necessary tools for the proper assimilation of the text. It was the author's purpose to make available, within one volume, the necessary material for a complete understanding of the radio phenomena with which the practical radio engineer must work.

The numerous elementary radio textbooks extant today are suitable for the beginner, but do not provide an adequate theoretical grounding for the serious student. On the other hand, the engineering texts, although many are excellent, are unnecessarily advanced for the student, for example, who is preparing himself for an FCC license examination. The author has attempted in this textbook to achieve a desirable medium between the popular operator's instruction book and the advanced engineering text. At the same time, the method of treatment is such as to preserve maximum utility as a reference and textbook for the engineer.

The aspirant for an FCC license will find in this book the necessary information to qualify him to pass any Federal Communications Commission examination. Readers in this category are urged to make use of the questions at the end of each chapter. These questions are taken from the FCC "Study Guide" from which the actual Federal examination questions are selected. For such students, it is suggested that the chapters be studied in the chronological order in which they appear in the book, and that the chapter on mathematics be thoroughly digested before proceeding with circuit theory.

It is suggested that the engineering student immediately proceed with the advanced alternating-current theory and electronic circuit theory. The preceding sections on mathematics, direct and alternating currents, inductance, and capacitance can then serve as "refresher" courses, as well as continuous reference material as the reader progresses through the text.

The author wishes to acknowledge the courtesy of *Radio News* in granting permission to use portions of an article written by the author which originally appeared in that magazine. Indebtedness is also gratefully acknowledged to the Astatic Corporation, the Blaw-Knox Company, the Brush Development Company, the Clarostat Manufacturing Company, Collins Radio Company, Communications Company, Inc., Cornell-Dubilier Electric Corporation, the Electric Storage Battery Company, Finch Telecommunications, Inc., the General Electric Company, General Radio Company, the Hammarlund Manufacturing Company, Inc., the International Derrick and Equipment Division of the International-Stacey Corporation. the International Resistance Company, Lapp Insulator Company, Inc., Lelugh Structural Steel Company, P. R. Mallory and Company, Inc., RCA Manufacturing Company, Inc., Radio Engineering Laboratories, Inc., Radio Receptor Company, Inc., Solar Manufacturing Corporation, Standard Transformer Corporation, Thomas A. Edison, Inc., Thordarson Electric Manufacturing Company, Truscon Steel Company, and the Western Electric Company for photographs, circuit diagrams, and data on commercial equipment.

The author also acknowledges the valuable suggestions of Captain Joseph R. Redman, U.S.N., Director Naval Communications, and Lieutenant Colonel Jay D. B Lattin, U. S. Army Signal Corps, Office of the Chief Signal Officer War Department, regarding subject matter to be included in the text. Finally, the author wishes to express his special appreciation to his wife, Margaret Haley Vogt, for her untiring assistance in typing the manuscript and reading proof, and for her never-failing encouragement without which the manuscript would unquestionably not have been completed.

ERNEST J. VOGT

# CONTENTS

# RADIO TECHNOLOGY

# Chapter I

# BASIC ESSENTIALS

The entire universe, of which the earth and solar system form so small a part, is composed of certain basic substances. There are more than ninety of these substances as found in nature. They are called **elements,** since they are the elementary materials of which everything in creation is composed. Everything that exists - the air that is inhaled, the house that is inhabited, the earth itself trees, plants, flowers— every animate and inanimate thing is composed of these fundamental elements. In many cases the elements exist alone; in others they combine with other elements to form more complex substances. Two of the familiar elements found in air are nitrogen and oxygen. Copper, aluminum, and nickel are three well-known elements that find special uses in radio.

## *THE ATOMIC THEORY*

**The Atom.** All elements exist in the form of extremely minute particles called **atoms.** An atom, in turn, is composed of a number of positive charges of electricity called **protons,** negative charges called **electrons,** and neutral entities called **neutrons.** The neutrons are sometimes called **nuclear electrons,** as distinguished from the negative electrons, which are called **orbital electrons.** The nomenclature of protons, neutrons, and electrons will be adhered to in this book. The sole difference between atoms of different elements lies in the number of protons, neutrons, and electrons in the atoms. Thus, an atom of carbon contains six electrons; an atom of oxygen contains eight electrons If two electrons could be added to the atom of carbon and it could be arranged correspondingly to compensate the proton charges to maintain atomic equilibrium, the atom of carbon would change to oxygen.

The electrons within an atom are in a state of constant motion, and according to the present theory they whirl at tremendous speeds in orbits about a center nucleus. The center nucleus is composed of a concentration of the positive charges, or protons, separated from each other by neutrons. The latter act as compensating forces to prevent the repellent tendencies of the protons from causing them to fly apart. The action of the electrons whirling about the nucleus can be compared to the movement of the planets about the sun.

**The Molecule.** Most of the forms of matter with which we come in

contact in our everyday experience are more or less complex combinations of the atoms of various elements. These combinations of atoms are called **molecules.** Ordinary water in its pure form, for example, is composed of countless molecules containing atoms of hydrogen and oxygen. In this case the proportion is two atoms of hydrogen to one of oxygen in each molecule, from which the familiar notation $H_2O$ is deduced. The molecules of any substance are in a constant state of movement, or

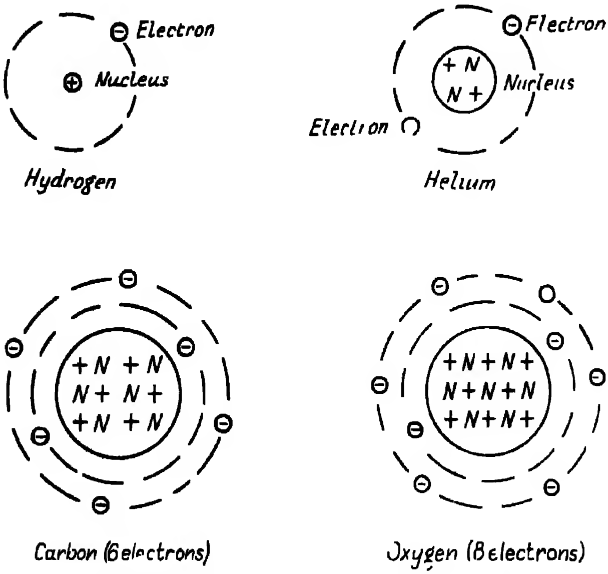



FIG 1. Schematic representation of different atoms , indicates a proton, indicates an electron, and $N$ indicates a neutron

agitation, owing to the thermal energy possessed by the substance. In a gas, the molecules possess a haphazard movement throughout the confines of the container, whereas, in a solid, only vibratory movements in relatively small regions are possible

Some idea of the relative size of electrons, atoms, and molecules can be obtained from the following analogy. If a drop of water were magnified to 1,000 times the size of the earth, a single molecule of water would be about a half mile in diameter. The individual atoms of hydrogen and oxygen would be about a quarter of a mile in diameter. The individual electrons would be about as large as a pea. It is apparent, therefore, that the electrons within an atom are by no means closely packed. Where the electrons are loosely bound to the parent atom, as in metals, the electrons can travel through the metal quite readily if they are provided with the proper stimulus.

## *ELECTRICITY*

**Electric Charges.** The ancient Greeks early discovered that if amber were rubbed upon cat's fur, the amber acquired the property of attracting bits of paper or pith, and they called this phenomenon **electrification,** after the Greek word for amber. In later years, considerable experimentation was carried on along these lines, leading to many more or less accidental discoveries. For example, it was discovered that electricity generated in this manner could be transmitted to other articles. Thus, it was found that a pith ball after being attracted to an electrified piece of amber, itself became electrified and could, in turn, attract other articles. It was found, too, that electricity could be generated in other ways. A glass rod that had been rubbed with a piece of silk was found to have become electrified. The electrified glass rod, in turn, was able to electrify a pith ball. Here, however, a startling phenomenon became evident. A pith ball that had been electrified from amber and another one electrified from a glass rod exerted strong attraction for each other. When placed in proximity to each other, both balls immediately flew together. On the other hand, two pith balls that had both been electrified from a piece of amber repelled each other. Similarly, two pith balls electrified from a glass rod flew apart when placed in proximity to each other. The two different types of electrification were called **positive** and **negative charges.** We have the ancient Greeks to thank, therefore, for our first fundamental law of electricity: *Like charges of electricity repel each other; whereas unlike charges attract each other.*

**Electric Currents.** Subsequent investigation led to the hypothesis that a negatively charged body was occasioned by an excess of electrons within the atoms of that substance, and, conversely, that a positively charged body was occasioned by a deficiency of electrons. In other words, in a negatively charged body, there are more electrons than protons, and a negative charge prevails. In a positively charged body, there are more protons than electrons, and a positive charge prevails. If a positively charged body and a negatively charged body are connected (by the physical insertion of a piece of metal, for example), we find that in accordance with our fundamental law these unlike charges attract each other. Electrons flow from the negatively charged body to the positively charged body and continue to flow until a state of electrical balance is reached. In other words, if the original positive and negative charges were of equal magnitude, both bodies have become neutralized, and there no longer exists any deficiency or excess of electrons. By performing such an experiment, we caused electrons to flow through the piece of metal used to connect the two bodies. This flow of electrons is called **electric current flow.**

From the foregoing discussion it is apparent that an electric current

refers to the movement of electrons in a substance. However, since electrons are normally in motion within the atom, we must further qualify this statement. The movement of electrons must refer to the movement of electrons *between* atoms. This so-called movement does not imply that a specific electron leaves the negatively charged body and then skips from atom to atom until finally it reaches the positively charged body. On the contrary, the positively charged body exerts a strong attraction upon the electrons of the atoms immediately adjacent to it. The negatively charged body also exerts a strong attraction upon the nuclei of the atoms adjacent to it (see Fig 2). This action is carried on successively through the atoms of the conducting substance until the effect is felt throughout the conductor. The speed of this effective

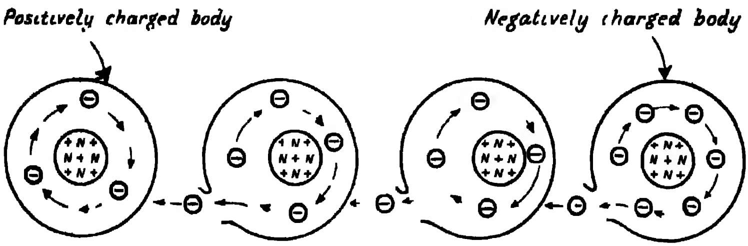



FIG 2 Movement of electrons in a conductor Notice the excess of electrons in the negatively charged body and the deficiency in the positively charged body The positive nuclei of the conducting atoms, it will be noted, have been displaced slightly from center due to the attraction of the negatively charged body

electron flow is equal to the speed of light - 186,000 miles per second. The rate of flow does not refer to the velocity of the actual electrons but to the rapidity with which a positive charge makes itself felt through a conductor to a negatively charged body at the other end of the conductor

The above phenomenon can be somewhat clarified by means of a mechanical analogy. We have all observed at one time or another a long freight train starting from a standstill and the delayed movement of the cars as the slack is taken up in the couplings between the cars In the case of an extremely long train having one locomotive pulling and one at the opposite end of the train pushing, if both locomotives started moving at precisely the same instant, the action could be compared to that of an electric charge. The movements of the pusher locomotive would send a series of bumps along the train as the slack in the couplings between cars was taken up. Similarly, the pulling locomotive would transmit a series of tugs along the train as the slack in the couplings between cars was taken up in the opposite direction. If both locomotives have the same power and start with the same speed, the tugs and bumps will meet in the center of the train, and at the instant of meeting, the

entire train will be in motion. The actual speed of the train, once it is moving, can be compared to the normal velocity of electrons within the atoms of a substance. It can be seen, however, that an appreciable interval of time elapsed between the initial tugs and bumps of the respective locomotives and the final movement of the *entire* train in response to them. The rapidity with which these tugs and bumps were transmitted over the length of the train until they were felt by the center car can be compared to the velocity of an electric charge through a substance. The locomotives can be compared to the positive and negative charges applied to the substance. In the train illustration, the time interval was an appreciable one. The velocity of an electric charge, however, is so phenomenally high that for all practical purposes the action can be considered instantaneous.

**Insulators and Conductors.** In the illustrations just cited, an electric current was caused to flow by connecting two oppositely charged bodies with a piece of *metal*. Had these bodies been connected with a piece of *porcelain*, very little current would have flowed. It would seem, therefore, that there are characteristics other than the normal atomic structure of substances to be taken into account when considering electric current flow through these substances. It has been found that the atoms of certain substances have very little inclination to release an electron under the attraction of a positive charge. Such materials as glass, porcelain, mica, and, to a lesser extent, Bakelite, rubber, cambric, fabrics of various kinds and paper fall into this classification. Such substances are **insulators**, since they may effectively be used wherever it is desired to *prevent* flow of electric current. On the other hand, the atoms of other substances very readily relinquish an electron under the attraction of a positive charge. Such substances, notably the metals, are called **conductors** and are used wherever it is desired to conduct a flow of electric current with a minimum of opposition. Copper is the metal most commonly used as a conductor, since it combines high conductivity with comparatively low cost. Silver is an even better conductor than copper, but the cost makes extensive use of it prohibitive. Other metals used as conductors for specific applications are, in the order of their conductivity, aluminum, zinc, iron, nickel, steel, and brass.

**Resistance.** No material is a perfect conductor, and no material is a perfect insulator. There is no sharp distinction between conductors and insulators. A substance that in some instances would be regarded as an insulator would, in other circumstances, be regarded as a conductor; for example, a material that is a good insulator at low temperatures may be a fairly effective conductor at high temperatures. Materials simply vary in the amount of opposition they offer to electric-current flow. This opposition to electric current flow is called "resistance," and the unit of resistance is called the **ohm.** The international standard ohm was first

defined in 1894 as follows: *One ohm is the resistance offered by a column of mercury* 106.3 cms *long having a cross-sectional area of* 1 sq mm. The ohm is merely an arbitrary value that has been accepted as standard in order that the resistance of any material in relation to this value may be determined. Thus, a piece of No. 14 B. & S. gauge copper wire 380 ft long has an approximate resistance of 1 ohm; a piece of No. 10 B. & S. gauge copper wire 1,000 ft long has a resistance of approximately 1 ohm.

The resistance of conductors remains comparatively constant under unvarying conditions. The resistance *does* vary, however, with a variation of temperature. Generally speaking, the resistance of all pure metals increases as the temperature increases. The increase in resistance (in ohms) for each degree rise in temperature is called **temperature coefficient of resistance.** The temperature coefficient of an alloy is generally less than the average of the coefficients of its constituents, whereas the resistance of certain alloys does not increase at all with an increase of temperature.

Naturally, the resistance of any given conductor will be a direct function of its physical dimensions. The resistance of a given piece of copper wire, for example, will be directly proportional to its length, but it will also vary with the thickness of the wire. Specifically, the resistance of a wire varies inversely with the area of its cross section.

**Conductance** is a measure of the conductivity of a material. Numerically, it is the reciprocal of the resistance. Thus,

$$G = \frac{1}{R}, \tag{1}$$

where $G$ = standard symbol for conductance;
$R$ = standard symbol for resistance.

Since conductance is the opposite of resistance, the unit of conductance has been made the mho (ohm spelled backward). Thus,

$$G\text{ (in mhos)} = \frac{1}{R\text{ (in ohms)}}, \tag{2}$$

and

$$R\text{ (in ohms)} = \frac{1}{G\text{ (in mhos)}}. \tag{3}$$

**Breakdown of an Insulator.** Electronically, there is no fundamental structural difference between conductors and insulators. Rather, the distinction lies in the relative cohesion of the electrons and protons within the atoms (refer to Fig. 3). A good insulator, when used within its limits, will not release *any* of the electrons from its atoms under the influence of an electric charge. Nevertheless, an electric charge has a definite effect upon these electrons and also upon the protons of the

center nucleus. The electrons are drawn toward the positive charge and the protons toward the negative charge, causing a *distortion* of the atom. For every type of substance, if the electric charge is made great enough, a point is reached where this distortion is so great that an electron finally does break away from the atom. When this happens, we say that the insulator has been "ruptured," and partial breakdown has occurred. Actually, when partial breakdown occurs, it is usually rapidly followed by complete breakdown. The freed electron bumps into other severely strained atoms, knocking other electrons loose. The action is cumulative

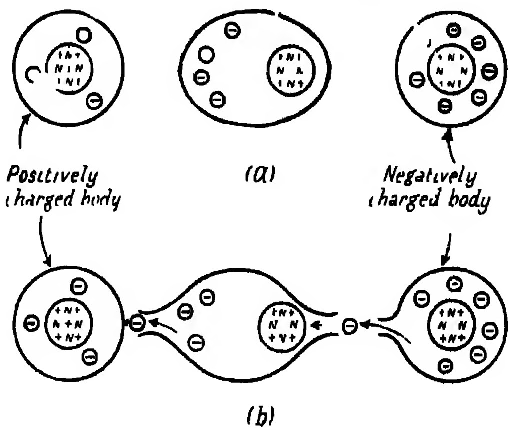



FIG. 3 An insulator atom under the influence of an electric field (a) Distortion, (b) Breakdown, or rupture

and results in complete disintegration of the insulator, which then becomes a conductor at the point of breakdown.

The effect of temperature upon insulators is very important. Any increase in temperature increases the agitation of the molecules and atoms of a substance in their motion, which in turn increases the frequency and violence of the collisions between atoms and between molecules. If the electrons are already straining to break away from their atoms because of the influence of an electric charge, such collisions will tend further to weaken the atomic structure and cause electrons to be released. It can be seen, therefore, that the efficiency of an insulator decreases as the temperature increases.

## MAGNETISM

Everyone is familiar with the action of the common magnet. If a piece of iron is placed in proximity to a magnet, it will be attracted to the magnet, and the magnet is said to be surrounded by a **magnetic field** and that any iron within the confines of this field will feel the attraction

of the magnet. Interesting as this phenomenon is historically, its use in modern engineering is comparatively limited.

**Electromagnetism.** In an earlier part of this chapter we discussed the effect of connecting a positively and a negatively charged body with a conductor. When a conductor has thus been placed within range of the attractive and repellent forces of such bodies, it is said to be in the presence of an **electric field.** Placing a conductor between such bodies, we have seen, results in a flow of current in the conductor. The electric field is present wherever there are two oppositely charged bodies, whether

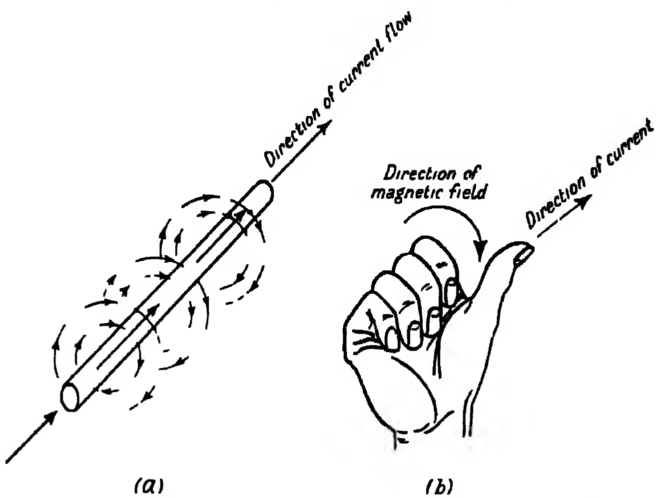



FIG 4. (a) The field about a current-carrying conductor (b) Illustrating the right-hand rule.

there is a conductor or not If a conductor is connected in this way, however, current flows, and the phenomenon of a *moving* electric field is present. It has been determined experimentally that such a moving field has the peculiar effect of creating a magnetic field. Such a magnetic field has the identical properties of the magnetic field surrounding a natural magnet. It has also been demonstrated that, conversely, a moving magnetic field causes an electric field. As a result of these experiments, the two basic laws were formulated which produced the groundwork for much of our present-day electrical industry. These laws are:

1. *A moving electric field creates a magnetic field.*
2. *A moving magnetic field creates an electric field.*

An electromagnetic field is always at right angles to the direction of motion of the electric field that produces it (see Fig. 4). In general, the strength of an electromagnetic field is directly proportional to the strength

of the current. The magnetic field about any given conductor can be greatly increased in intensity by winding the conductor in the form of a coil, as in Fig. 5. The fields about adjacent turns add to each other, resulting in a comparatively powerful field concentrated in a smaller area than would be the case if the conductor were stretched out in a straight line. The actual strength of the field, of course, will be a direct function of the current intensity and the number of turns in the coil.

The direction of the magnetic field about a conductor can be easily determined by the application of the *right-hand rule* (see Fig. 4). If the conductor is grasped in the right hand with the fingers around the conductor and the extended thumb pointing in the direction of the current

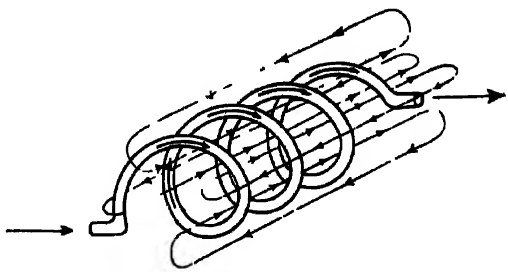


FIG. 5 Magnetic field about a coil Notice how the lines of force, instead of revolving about the individual turns add to the fields of adjacent turns, creating a longer and more powerful field

through the conductor, the fingers will point in the direction of the magnetic field.

**Magnetic Flux.** From the right-hand rule it is apparent that a magnetic field about a conductor follows a very definite pattern. This pattern is composed of a number of imaginary lines along which the attractive or repulsive force of a magnetic field acts. Technically, they are known as **lines of force,** sometimes called "lines of magnetic induction." The total number of lines comprising a given magnetic field is known as the **flux** of the magnetic field. The density of the flux of a particular field depends upon the nature of the medium through which it is passing. If a certain part of this medium allows a freer passage of the lines of force through it than does the surrounding medium, the major portion of the total flux will be concentrated in that part of the medium. The **flux density** can be defined, therefore, as the number of lines of force per unit area passing through any substance through a plane at right angles to the direction of the lines of force.

**Magnetic Permeability.** Materials offer varying opposition to the passage of lines of force. The ability of a material to facilitate the passage of these flux lines is known as its **magnetic permeability.** The magnetic permeability of air is taken as a reference point and is assigned a value

of unity. The *specific magnetic permeability* of a substance may be defined as the ratio of the flux density produced in this substance by a certain magnetizing force to the flux density produced in air by the same magnetizing force. Most substances have a permeability of approximately unity. The most notable exceptions are nickel, cobalt, and iron, with iron having exceptionally high magnetic permeability. The permeability of a good grade of magnetic iron may be as high as several thousand.

## *TERMINOLOGY*

**The Coulomb.** In order to proceed with the practical application of electric currents, it is necessary to assign quantitative values so that we may specifically refer to such currents. Since the smallest quantity of electricity is the electron, it is fitting that electricity should be measured in terms of the electron. The standard unit of *quantity* is the **coulomb.** It has been named in honor of the French physicist Charles Augustin Coulomb in recognition of his many contributions to the science of electricity in the late eighteenth and early nineteenth centuries. A coulomb is the amount of electricity represented by 6.28 times $10^{18}$ electrons.

**The Ampere.** In dealing with electric currents, it has become necessary to find some designation with which to represent the amount of electricity flowing past a point in a conductor during a given unit of time. This unit designation for current is called the **ampere.** Since one of the earliest uses of electricity was in electroplating, one ampere was designated as that flow of electrons which will deposit 1.118 milligrams of silver per second from a silver nitrate solution in a standard voltameter. The ampere was named after another French scientist, André-Marie Ampère, who lived in the early nineteenth century.

The ampere and the coulomb should not be confused. The coulomb represents *quantity*; the ampere, *current*. In this case, an electric current might be compared to water flowing through a pipe. One ampere of electricity flowing through a conductor is analogous to, for example, one gallon of water *per second* flowing past a certain point in the pipe. A coulomb is the quantity of electricity transported by a current of one ampere flowing for one second.

**The Volt.** Since the positively and negatively charged bodies responsible for the flow of current in a conductor represent a *pressure*, or *electromotive force*, it is necessary that a unit be designated to enable us quantitatively to measure this force. The unit of electromotive force, abbreviated emf, is the **volt,** named in honor of the Italian physicist Alessandro Volta. The specific value of the volt has been standardized by modern physicists as 1/1.0183 of the electromotive force generated by a standard Weston cell. In other words, a standard Weston cell generates 1.0183 v.

Continuing the previous hydraulic analogy, the volt, or voltage, may be compared to the pressure in a system of water pipes. When it is said that a certain number of pounds per square inch pressure will cause so many gallons of water per second to flow through a certain pipe, it is analogous to saying that a certain voltage will cause a resultant current of so many amperes in a certain conductor. The ability of a conductor to present a resistance to a flow of current is measured in terms of another unit, the ohm. One volt is the electromotive force necessary to produce a current flow of one ampere against a resistance of one ohm. This is discussed further in Chap. IV.

**The Joule.** Where positive and negative electricity are separated from each other, as they are in batteries and generators, these charges possess potential energy. If appropriate circuits are established through which these charges can recombine, the potential energy possessed by the charges can be changed to heat, light, mechanical energy, or even chemical energy. Hence, electricity is a physical agent which is capable of doing work. As such, some means had to be devised to designate the amount of work done by this energy in order to correlate it with work accomplished by other means. The unit of work is the **joule.** One joule is the amount of work done in one second by a current of one ampere at a pressure of one volt. The joule was named after the British scientist James Prescott Joule, who was famed for his researches into the theory of the conservation of energy.

The joule is directly related to other units of energy as follows:

$$
\begin{aligned}
1 \text{ joule} &= 6.25 \times 10^{18} \text{ electron-volts,} \\
&= 0.238 \text{ gram-calorie,} \\
&= 0.738 \text{ foot-pound,} \\
&= 10{,}000{,}000 \text{ ergs.}
\end{aligned}
$$

**The Watt.** The unit of electrical power is the **watt.** One watt is the rate at which work (in joules per second) is done by a current of one ampere at a pressure of one volt. The watt should not be confused with the joule. The joule represents the *quantity* of work done; the watt represents the *rate* at which this work is done.

$$P = \frac{W}{t}, \tag{4}$$

where $P$ = power in watts;
$W$ = amount of work done in joules;
$t$ = time in seconds in which $W$ is done.

It is apparent, therefore, that 1 watt is the rate of work represented by 1 joule per second; 746 watts equal 1 horsepower. The watt was named after the Scotsman James Watt, famous for his work on power, especially in the field of steam.

**Electrical Prefixes.** It has become necessary in dealing with electrical quantities to adopt units which are multiples or submultiples of the standard units. This greatly facilitates the handling of extremely large or extremely small quantities of electricity. Instead of coining a redundancy of additional names for these multiple units, a system was

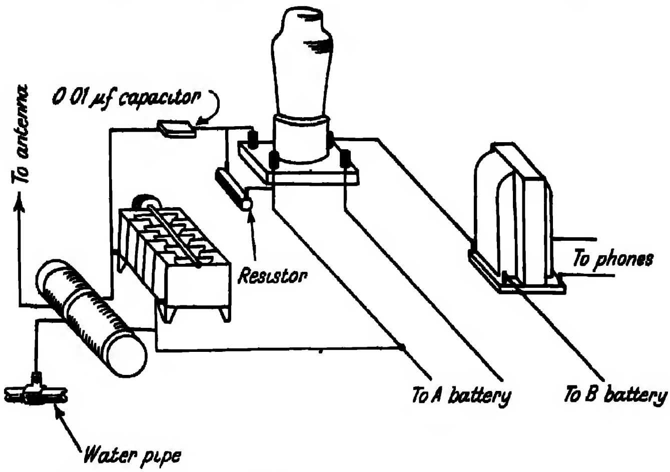



FIG. 6 Pictorial representation of a simple receiving circuit

devised utilizing the principle of the metric system. The following is a tabulated list of the standard prefixes.

| *Prefix* | *Meaning* |
|---|---|
| micromicro- | one million millionth |
| micro | one millionth |
| milli | one thousandth |
| deci | one tenth |
| kilo- | one thousand |
| mega | one million |

These prefixes are equally applicable to any of the standard units. Thus, one milliampere is a thousandth of an ampere, one millivolt is a thousandth of a volt; one kilovolt is a thousand volts, one megohm is a million ohms. There are many standard units that the reader will encounter as he progresses through the text, and the above prefixes are applicable to nearly all.

In the following chapter, which deals with mathematics, a section is devoted to the use of exponents for the facile conversion of units to the various prefixes. The exponent system enables the user to avoid the handling of large, cumbersome figures in making conversions.

**Symbols.** In radio, as in any exact science, clear thinking occurs only when one has a precise language that is *accurately expressed.* The expression of the science of radio by word is not sufficiently complete to be

all-inclusive, despite the fact that no vocabulary is more precise than that used in radio. In order to exchange ideas efficiently, it is necessary to supplement word description with illustration.

By illustration we do not mean mere pictorial representation. To the layman, the circuit illustrated in Fig. 6 may seem eminently satisfactory. To the technician or advanced student, the drawing is comparatively meaningless. Figure 6 is an illustration of a one-tube radio receiver. As a means of conveying information concerning this receiver, the illustration is extremely unsuitable. What is the relation of primary to secondary in the r-f coil? Is the relative movement along the axis? Or at an angle? What type of tube is being used? Is it a pentode? A tetrode? A triode? Is the filament directly or indirectly heated? Does the output transformer have a core? (We may assume so from the manner in which it is used.) Is the core shown in the illustration?

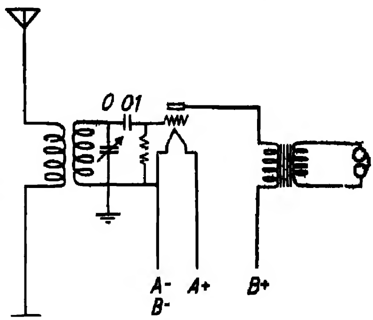


FIG. 7. Schematic diagram of the circuit shown pictorially in Fig. 6.

Figure 7 illustrates a *schematic* diagram of the same circuit. Notice first its extreme simplicity. Every needed factor is shown. The functions of the circuit can be read at a glance. No cumbersome, time-consuming artistry is attempted. The fixed capacitor is efficiently depicted by two parallel lines; the variable capacitor by two parallel lines with an arrow drawn through them. In the pictorial diagram, a fixed capacitor might appear in a variety of shapes or forms, and the only clue to its identity would be a sign labeling it as such. The only information required about a fixed capacitor in a circuit is its value and its use in the circuit. Both can be adequately represented by two parallel lines with the proper notation.

The symbols for the component parts that make up schematic diagrams are many and varied. A complete list of standard symbols used in radio is given in Table I in the Appendix. The student is urged to memorize them and to refer to them as often as he encounters diagrams throughout the text.

In general, there are two types of diagrams utilized in radio work, namely, the *schematic* diagram and the *wiring* diagram. The schematic diagram is used wherever information is sought regarding the principle of operation, the theory, or the actual mechanics of a circuit. The wiring diagram is used whenever component parts or actual wires must be located on a specific piece of apparatus. Both types of diagrams have definite uses. Figure 8 is a wiring diagram of the receiver shown schematically in Fig. 7 as it might appear for a given chassis.

Suppose, for example, that this particular receiver becomes inoperative and the cause of the trouble is being ascertained. From routine tests, it is deduced that the trouble lies in an open grid circuit. Reference to the schematic diagram discloses that such an open circuit can occur only in the secondary of the r-f transformer, in the grid capacitor, or in the wires connecting these components to each other and to the tube. The next step, obviously, is to test these components and wires in order to find which of them is open. The *wiring* diagram is now consulted. From it we learn the exact location on the chassis of each of the component parts and the wires connecting them, thus greatly facilitating the service work.

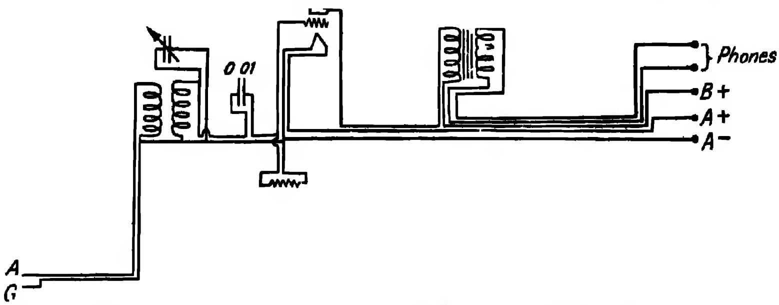



FIG. 8. Wiring diagram of the circuit of Fig. 6 as it might appear for a given chassis.

Of course, this receiver is a very simple circuit, and the above procedure would undoubtedly be unnecessary. In a complicated circuit such as a modern transmitter, however, the proper use of wiring and schematic diagrams saves a great deal of labor and time.

## QUESTIONS AND PROBLEMS*

1. If the diameter of a conductor of given length is doubled, how will the resistance be affected?
2. Name four conducting materials in the order of their conductivity.
3. Define the term "permeability."
4. Define a negatively charged body. A positively charged body.
5. What factors influence the direction of magnetic lines of force generated by an electromagnet?
6. What is the unit of resistance?
7. What is the unit of electrical power?
8. What is the unit of conductance?
9. Define the term "coulomb."
10. What effect does the cross-sectional area of a conductor have upon its resistance per unit length?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

# Chapter II

# RADIO MATHEMATICS

The science of algebra is based on arithmetic. The fundamental operations of algebra, just as in arithmetic are addition, subtraction, multiplication, and division. In fact, there is no clear line of demarcation between arithmetic and algebra. The fundamental principles of each are identical. In algebra, however, the objective is to generalize the truths learned in the study of arithmetic. Thus, in arithmetic one deals with the Arabic *numerals* representing specific quantities. In addition to numerals, algebra deals with *letters*, which may represent known and unknown quantities. The use of algebra, therefore, permits the application of the principles of arithmetic to a much broader field.

## *ELEMENTARY ALGEBRA*

**Positive and Negative Numbers.** In arithmetic, the signs + and − are used to indicate addition and subtraction. In other words, these signs are used to indicate the operation to be performed. In algebra, in addition to the above usage, these signs are used to indicate opposite *states* of quantities; a number or quantity must be either a *positive* or a *negative* quantity. The absence of any sign preceding a number is taken to indicate that the quantity is positive.

The concept of positive and negative quantities is derived from the fact that in the study of any science physical quantities often exist in opposite states. Thus, in radio, there are *losses* and *gains* in amplifiers, *amplification* and *degeneration*, there are *plus* voltages and *minus* voltages in a circuit. The handling of such opposite quantities in mathematical problems is greatly facilitated by representing them as + or − quantities. Thus, if a certain *gain* in an amplifier is represented as + 10 db, an equal *loss* in the amplifier must be shown as − 10 db.

In algebra therefore, it should be remembered that one may count, not only from zero *upward*, as in arithmetic: 0, 1, 2, 3, 4, 5, 6, 7—but also from zero *downward*. This can be represented on a horizontal scale as follows: − 9, − 8, − 7, − 6, − 5, − 4, − 3, − 2, − 1, 0, + 1, + 2, + 3, + 4, + 5, + 6, + 7, + 8, + 9, where all values to the right of zero are positive values and all values to the left of zero are negative values.

The operations of addition and subtraction are thus reduced to counting

along a scale of numbers. For example, 3 is added to 4 by beginning at $+4$ in the scale and counting 3 units in the ascending, or *additive*, direction, giving the result of $+7$. Three is subtracted from 4 by beginning at the $+4$ in the scale and counting 3 units in the descending, or *subtractive*, direction, giving the result of $+1$. By extending this rule to apply in all cases, it is possible to perform operations that have no counterpart in arithmetic. Thus, 7 can be subtracted from 4 by beginning at $+4$ in the scale and counting 7 units in the subtractive direction, obtaining the result $-3$.

The above operations of algebraic addition and subtraction may be generalized in three rules, which the student should memorize:

*Rule 1. When adding two numbers of like signs, add in the usual arithmetical manner, and the answer will take the sign of the original numbers.*

Thus,

$$+4 \text{ plus } +3 = +7. \tag{1}$$
$$-4 \text{ plus } -3 = -7. \tag{2}$$

*Rule 2. When adding two numbers of unlike signs, subtract the smaller from the larger, and the answer will take the sign of the larger.*

Thus,

$$+4 \text{ plus } -8 = -4. \tag{3}$$
$$-7 \text{ plus } +9 = +2 \tag{4}$$

*Rule 3. When subtracting one number from another, change the sign of the subtrahend and add.*

Thus,

$$4 \text{ minus } -6 = +10. \tag{5}$$
$$+10 \text{ minus } +2 = +8. \tag{6}$$

When adding or subtracting quantities involving more than two numbers, the general procedure is to gather all the numbers of like signs and add them. The indicated operation is then performed with the resultant quantities.

Thus, to add $+7$, $-3$, $-8$, $+14$, $+11$, and $-6$,

| | | |
|---|---|---|
| $+7$ | $-3$ | $+32$ |
| $+14$ | $-8$ | $-17$ |
| $+11$ | $-6$ | |
| | | $+15$ *Ans.* |
| $+32$ | $-17$ | |

The operations of multiplication and division are performed in algebra just as they are in arithmetic. In algebra, however, one must take cognizance of the positive and negative states of the quantities involved. The rules governing the sign of the product or quotient of these operations are as follows:

***Rule 4. When multiplying or dividing numbers having like signs, the answer is always positive.***

Thus,

$$4 \cdot 2 = +8, \tag{7}$$
$$-4 \times -2 = +8, \tag{8}$$
$$\frac{4}{2} = +2, \tag{9}$$
$$\frac{4}{-2} = +2. \tag{10}$$

***Rule 5. When multiplying or dividing numbers having unlike signs, the answer is always negative.***

Thus,

$$8 \times 2 \quad 16, \tag{11}$$
$$8 \quad 2 \quad 16. \tag{12}$$
$$\frac{8}{4} \quad 2, \tag{13}$$
$$\frac{8}{4} \quad 2. \tag{14}$$

In algebra, the multiplication sign ($\times$) is dispensed with because of the possibility of its being confused with the letter $x$. Multiplication is indicated simply by a dot ($\cdot$) between the two quantities involved. Thus, $2 \cdot 2$ means the same as $2 \times 2$. Where the quantities are expressed in letters, the multiplication sign is usually omitted. Thus, $IR$, $I \times R$, and $I \cdot R$ all have the same meaning.

Multiplication is also often indicated by enclosing one or both of the terms concerned in parentheses. Thus, $a(b)$, $(a)(b)$, and $ab$ have the same meaning. Similarly, $x(a + b)$ means that the term $a + b$ is to be multiplied by $x$. $x(a + b)$, therefore, is equal to $xa + xb$. Wherever two or more terms are combined, it is customary to arrange them in the progressive order of the alphabet. Thus, $xa + xb$ is more properly written $ax + bx$. This is simply a matter of form, since the numerical value of the expression is unaffected by the order in which it is written.

Parentheses are used whenever it is desired to indicate that a term is to be treated as a single-number expression. Any group of terms related to each other by addition or subtraction must be treated as a *single* term so far as any operation other than the indicated ones (addition or subtraction) is concerned. Thus, in the expression $a - b - c$, it is impossible to divide or multiply only the $b$ by a number. The entire term as a whole must be divided or multiplied by the number. Parentheses are used to indicate the operation to be performed in the case of multiplication. Thus, if it is desired to multiply $a - b - c$ by 2, the operation is indicated $2(a - b - c)$, and the result is $2a - 2b - 2c$.

Any one of the factors into which an indicated product can be dissolved is called the **coefficient** of the remaining factors. Thus, in the expression $abc$, $a$ is the coefficient of $bc$; $b$ is the coefficient of $ac$; and $c$ is the coefficient of $ab$. If an indicated product is composed of a numerical and a literal factor, it is customary to write the numerical factor, or coefficient, first. Thus, $5ab$ is approved form, but $a5b$ and $ab5$ are not. If a term is composed of a known and an unknown factor, both literal, it is customary to write the known factor, or the coefficient, first. For example, if $a$ is known, $x$ is unknown, and it is desired to indicate the multiplication of $a$ and $x$, the proper written form is $ax$. $xa$, although having the identical numerical value, is not approved form.

The algebraic signs for division remain the same as those utilized in arithmetic, although it is usually more convenient to express a division in the fractional form. Thus, $I \div R$, $I/R$, and $\frac{I}{R}$ have the same meaning.

In performing the operations of addition, subtraction, multiplication, and division with letters having unknown values, the operation is carried out as far as possible under conditions of the problem, and the result is then *indicated* in the simplest form to which it has been possible to reduce the problem. Thus, $a \cdot a = a^2$, just as $2 \cdot 2 = 2^2$ or 4. There is no possible way in which we may actually multiply $a$ by $a$ unless the numerical value of $a$ is known. However, we may *indicate* that $a$ has been multiplied by itself by use of the notation $a^2$. The product of $ab$, however, can be simplified no further, since $a$ and $b$ represent separate unknown quantities. Similarly, the final expression representing the division of $a$ by $b$ is $\frac{a}{b}$. In all other respects, letters in algebra can be handled precisely as are numbers. Thus, $a + a = 2a$; $b + b = 2b$; $\frac{a}{a} = 1$; $\frac{2b}{b} = 2$; and so on; since in any given problem the value of a given quantity remains the same.

**The Simple Equation.** An **equation** is simply a mathematical expression of an equality. Thus, the expression $4 + 3 = 7$ is an equation. An equation is said to consist of two *sides*, or *members*. In the above equation, $4 + 3$ is the **left member**, 7 is the **right member.**

Any number that satisfies the condition of an equation is called a **root** of the equation. Consider the equation

$$x + 4 = 10. \tag{15}$$

Since, by definition, an equation is an expression of an equality, it follows that $x$ in Eq. (15) must equal 6. Since 6 is the number which satisfies the condition of the equation, 6 is the root of the equation. Simple equations have but one root.

**Solving** an equation is the process of finding the root or roots of the

equation. The final objective in solving an equation, therefore, is to have the unknown isolated in one member of the equation and all the remaining values in the other member of the equation. In achieving this objective, *anything* may be done to an equation, provided that the operation yields an equivalent equation. **Equivalent equations** are equations having the same roots, or solutions.

The following rules will serve to guide the student in the solution of simple equations:

1. *The same expression may be added to both members of an equation.*
2. *The same expression may be subtracted from both members of an equation.*
3. *Both members of the equation may be multiplied or divided by the same expression, provided that this expression is not zero and does not involve the unknowns of the equation.*

It should be remembered that none of these operations may be performed on just *one* member of an equation. Whatever operation is performed on one member of an equation must also be performed on the remaining member. If this rule is not adhered to, the condition of equality is destroyed and the resulting expression is therefore no longer an equation.

The above rules have led to certain mechanical processes that considerably facilitate the solution of equations. Thus, a term appearing on both sides of an equation can be *canceled*.

$$x + a = 5 + a. \tag{16}$$

By cancellation

$$x = 5. \tag{17}$$

It is seen, therefore, that the process of cancellation is an application of Rule 2, since it is equivalent to subtracting the same term from both numbers.

A term can be *transposed* from one member to the other simply by changing its sign. Thus,

$$x + a = 5 + b. \tag{18}$$

By transposition,

$$x = 5 + b - a. \tag{19}$$

Similarly,

$$x + 5 = 7. \tag{20}$$

By transposition,

$$x = 7 - 5, \tag{21}$$

and

$$x = 2. \tag{22}$$

Transposition is simply another method of applying Rule 2 to the solution of equations. Thus, in Eq. (18), if $a$ is subtracted from both

members, the result is Eq. (19). Similarly, when 5 is subtracted from both sides of Eq. (20), the result is Eq. (21), which by arithmetic becomes (22).

Since the ultimate objective is to isolate the unknown in one member of the equation, the entire equation should be simplified as much as possible before attempting to segregate the unknown. The solution is therefore performed in progressive steps, each of which further simplifies the equation. The solution of a simple equation is best performed in the following steps:

1. Clear the equation of fractions by multiplying both members by the lowest common denominator of the denominators of the equation.
2. Remove all parentheses by performing the indicated operations.
3. Transpose all terms involving the unknown to the left member.
4. Transpose all the remaining terms to the right member.
5. Combine all terms in the unknown, exhibiting it as a factor.
6. Divide both sides by the coefficient of the unknown.
7. Check the solution by substitution in the original equation.

**Problem.** Solve

$$\frac{x-4}{3} - \frac{x-3}{2} - \frac{3+x}{10} = -2 \tag{23}$$

**Solution.** Clear of fractions by multiplying both members by the least common denominator 30. Thus,

$$10(x-4) - 15(x-3) - 3(3+x) = -60 \tag{24}$$

Remove the parentheses by performing the indicated multiplications

$$10x - 40 - 15x + 45 - 9 - 3x = -60. \tag{25}$$

Combine terms.

$$-5x + 5 - 3x = -51 \tag{26}$$

Transpose all terms involving $x$ to the left member

$$-5x - 3x + 5 = -51 \tag{27}$$

Combine terms

$$-8x + 5 = -51 \tag{28}$$

Transpose all terms *not* involving $x$ to the right member

$$-8x = -56 \tag{29}$$

Divide both sides by the coefficient of $x$, that is, by $-8$

$$x = 7. \tag{30}$$

Check by substitution in the original Eq. (23)

$$\frac{7-4}{3} - \frac{7-3}{2} - \frac{3+7}{10} = -2, \tag{31}$$

$$\frac{3}{3} - \frac{4}{2} - \frac{10}{10} = -2, \tag{32}$$

$$1 - 2 - 1 = -2, \tag{33}$$

$$-1 = -1.$$

**Ratio and Proportion.** The relation of two numbers that is expressed by the quotient of the first divided by the second is called the **ratio** of those numbers. The customary sign of a ratio is a colon (:). In radio work, it is usually more convenient to express a ratio in fractional form. Thus, the ratio of $a$ to $b$ may be written $a : b$, or $\frac{a}{b}$, or $a/b$.

An equality of two ratios is called a **proportion.** A fractional equation having one fraction in each member may therefore be called a "proportion." Thus, the equation

$$\frac{a}{b} = \frac{c}{d} \tag{34}$$

is a proportion. It can therefore be read: $a$ is to $b$ as $c$ is to $d$. This may be expressed: $a : b :: c : d$ where the double colon (::) is used in place of the sign of equality.

Proportions are often used in radio work, especially in circuits involving the Wheatstone bridge (see Chap. IV). A number of rules have been propounded governing the mathematical manipulation of proportions. Of these, two are used often enough in radio work to justify their discussion here The reader is urged to familiarize himself with them. They not only facilitate operations involving proportions but also often materially assist in the solution of equations involving fractions.

In a proportion, the first and fourth terms are the **extremes**; the second and third terms, the **means.** In Eq. (34), $a$ and $d$ are the extremes, and $b$ and $c$ are the means.

*Rule* 1. *In any proportion, the product of the means is equal to the product of the extremes.*

Given

$$\frac{a}{b} = \frac{c}{d}. \tag{34}$$

Then, according to Rule 1,

$$ad = bc. \tag{35}$$

This is readily proved by eliminating fractions from Eq. (34) by ordinary algebraic means, that is by multiplying both members by $bd$.

Rule 1 is often used in the solution of fractional equations that are proportions. It is almost always simpler and more rapid. Thus, in the equation

$$\frac{2}{7(a + x)} = \frac{3}{x + c} \tag{36}$$

it is much simpler to multiply the means and the extremes in order to clear of fractions, giving the result

$$21(a + x) = 2(x + c). \tag{37}$$

The alternative method of clearing fractions would involve the cumbersome least common denominator $7(a + x)(x + c)$.

When Rule 1 is applied to clear an equation of fractions, it is customary to say that the fractions have been cleared by **cross products.** This is derived from the fact that if a line is drawn through each pair of factors to be multiplied, the resulting configuration is a cross or $\times$.

*Rule 2. In any proportion, the ratio of the numerators is equal to the ratio of the denominators, that is, the terms are in proportion by alternation.*

Given

$$\frac{a}{b} = \frac{c}{d}. \tag{34}$$

Then, according to Rule 2,

$$\frac{a}{c} = \frac{b}{d}. \tag{36}$$

This is readily proved as follows: By cross products, Eq. (34) becomes

$$ad = bc. \tag{35}$$

Dividing both sides by $cd$, Eq. (35) becomes

$$\frac{a}{c} = \frac{b}{d}. \tag{36}$$

**The Simultaneous Simple Equation.** Two or more equations that are satisfied by the same set or sets of values of the unknown numbers form a system of **simultaneous equations.** For example, each of the equations

$$x + y = 5 \tag{39}$$

and

$$x - y = 1 \tag{40}$$

can be satisfied separately by an unlimited number of different sets of values for $x$ and $y$. Thus, Eq. (39) can be satisfied by

$$x = 4,\ y = 1,$$
$$x = 3,\ y = 2,$$
$$x = 2,\ y = 3,$$
$$x = 1,\ y = 4,$$

and so on. Similarly, Eq. (40) can be satisfied by

$$x = 2,\ y = 1,$$
$$x = 3,\ y = 2,$$
$$x = 4,\ y = 3,$$
$$x = 5,\ y = 4,$$

and so on. There is only *one* set of values, however, that satisfies the conditions of *both* equations, namely,

$$x = 3 \quad \text{and} \quad y = 2.$$

Since there *is* a set of values that satisfies both equations, Eqs. (39) and (40) are called **simultaneous equations.** They are also sometimes called **consistent equations.** When there is no set of values of the unknown terms that satisfies the conditions of a system of equations, the equations are called **inconsistent equations.** Thus, the equations

$$x + y = 7 \tag{41}$$

and

$$x + y = 9 \tag{42}$$

have no common set of values for $x$ and $y$ and therefore are inconsistent equations. Inconsistent equations cannot be solved by the methods applicable to the solution of simultaneous equations.

Simultaneous simple equations are solved by various processes of elimination, known as **elimination by addition or subtraction, elimination by comparison,** and **elimination by substitution.**

Elimination by addition or subtraction consists of eliminating the terms having the same unknown in both equations. This is accomplished by first multiplying (when necessary) one of the equations by a factor that will make the coefficient of one of the unknown terms equal to the term having the same unknown in the remaining equation. If these terms in the same unknown have *like* signs in each equation, they may be eliminated by *subtracting* the equations from each other. If they have *unlike* signs, they may be eliminated by *adding* the equations to each other. The resulting equation is then solved for the remaining unknown by the methods applicable to simple equations. When the equation is solved, the answer is substituted in one of the original equations, and this original equation is then solved as a simple equation to find the remaining unknown.

**Problem.** Solve

$$3x - 4y = 7, \tag{43}$$

$$x + 10y = 25. \tag{44}$$

**Solution.** Multiply Eq. (44) by the factor 3.

$$3x + 30y = 75. \tag{45}$$

Subtract Eq. (45) from Eq. (43), thus eliminating the terms involving the unknown $x$.

$$3x - 4y = 7 \tag{43}$$

$$3x + 30y = 75 \tag{45}$$

$$-34y = -68 \tag{46}$$

Then

$$y = \frac{-68}{-34} \tag{47}$$

and

$$y = 2. \tag{48}$$

Substitute the value of $y$ obtained in Eq. (48) in Eq. (43). This yields the equation

$$3x - 8 = 7. \tag{49}$$

Solving,

$$3x = 15, \tag{50}$$

$$x = \frac{15}{3}, \tag{51}$$

and

$$x = 5. \tag{52}$$

The solution that satisfies both of the original Eqs. (43) and (44) is $x = 5$ and $y = 2$. The answer can be checked by substituting these values of $x$ and $y$ in each of the original equations.

Elimination by comparison consists of finding an expression for the value of the same unknown term in each equation and then equating, or *comparing*, the two expressions. The equation thus formed is solved for its unknown. The answer is substituted in one of the original equations, which is then solved for the remaining unknown.

**Problem.** Solve

$$3x - 2y = 10, \tag{53}$$

$$x + y = 70. \tag{54}$$

**Solution.** The same unknown term is obtained in each equation by multiplying Eq. (54) by the factor 2. Equation (54) then becomes

$$2x + 2y = 140. \tag{55}$$

An expression for the same unknown term $2y$ in each equation is then found by solving each equation for $2y$ in terms of the remaining values. Equation (53) thus becomes

$$2y = 3x - 10. \tag{56}$$

Equation (55) becomes

$$2y = 140 - 2x. \tag{57}$$

Since the left members of both Eqs. (56) and (57) are equal to each other ($2y = 2y$), it follows that the right members must be equal to each other. This is derived from the fundamental mathematical axiom that *things equal to the same things are equal to each other*. The right members of Eqs. (56) and (57) may therefore be equated as follows:

$$3x - 10 = 140 - 2x. \tag{58}$$

Equation (58) is solved for its unknown, with the result

$$x = 30. \tag{59}$$

The value of $x$ obtained in Eq. (59) is then substituted in either of the original Eqs. (53) or (54). By substitution, Eq. (53) becomes

$$90 - 2y = 10. \tag{60}$$

Equation (60) is then solved for $y$, with the result

$$y = 40. \tag{61}$$

The solution that satisfies both of the original Eqs. (53) and (54) is therefore $x = 30$ and $y = 40$. The solution can be verified by substituting these values of $x$ and $y$ in each of the original equations.

Elimination by substitution consists of solving one of the equations given for one of its unknowns in terms of the remaining unknown. The answer thus obtained is substituted in the other equation which is solved for the remaining unknown. The value obtained for this unknown is substituted in either of the original equations to find the numerical value of the first unknown.

**Problem.** Solve

$$x - y = 4, \tag{62}$$
$$4y - x = 14. \tag{63}$$

**Solution.** Equation (62) is solved for $x$ in terms of the remaining values; thus,

$$x = y + 4 \tag{64}$$

This value of $x$, that is $(y + 4)$, is substituted in Eq. (63).

$$4y - y - 4 = 14 \tag{65}$$

Combining terms,

$$3y - 4 = 14 \tag{66}$$

Solving,

$$y = 6 \tag{67}$$

The value of $y$ obtained in Eq. (67) is then substituted in (62), which becomes

$$x - 6 = 4 \tag{68}$$

Solving,

$$x = 10 \tag{69}$$

The three standard methods of solving simple simultaneous equations by elimination are applicable under all circumstances. Nevertheless, each method has its advantages for special cases. The student should study each problem and select the method best suited for its solution.

**The Theory of Exponents.** A number or letter placed a little above and to the right of a number is called an **exponent** of the power thus indicated. The exponent indicates how many times the number is to be used as a factor, that is, how many times the number is to be multiplied by itself. Thus, in the expression $5^3$, 3 is the exponent. This indicates that 5 is to be multiplied by itself 3 times. $5^3$, therefore, is equal to $5 \cdot 5 \cdot 5$, or 125. The expression $5^3$ is read "5 to the third power" or "5 cubed."

Exponents find their greatest application in radio work in performing arithmetical operations involving extremely small decimal fractions or extremely large whole numbers. The solution of problems involving arithmetical operations with such cumbersome figures is greatly facilitated by converting these figures to whole numbers and *like* numbers having

exponents. *Like* numbers having exponents may very easily be combined in any arithmetical operation. *Unlike* numbers having exponents must first have the operation indicated by the exponent performed before they can be combined in any manner; for example, $8^5$, $7^3$, and $6^7$ are *unlike* numbers having exponents. Before any arithmetical operation can be performed with the three latter numbers, they must first be raised to the power indicated by the exponent. Since they offer no practical advantage as a short cut in radio work, *unlike* numbers having exponents will not be discussed here.

The main advantage of converting unwieldy decimal or whole numbers to whole numbers combined with like numbers having exponents is that arithmetical operations may be performed more quickly and easily and with fewer opportunities to make errors. The conversion to numbers having exponents is most easily accomplished by using exponents to the base 10, that is, by converting the number to a smaller whole number multiplied by 10 raised to the proper exponential power, because the conversion in this case consists simply of moving the decimal point of the number either to the right or to the left, as the occasion warrants. For this reason, nearly all the exponential numbers in radio problems are to the base 10.

The following rules govern the use of *like* numbers having exponents

*Rule* 1. *Two or more like numbers having exponents may be multiplied by adding the exponents and affixing the resulting exponent to the common number.*

Thus,

$$10^4 \cdot 10^2 \cdot 10^3 = 10^9, \tag{70}$$
$$10^2 \cdot 10^3 \cdot 10^6 = 10^{11}, \tag{71}$$
$$8^6 \cdot 8^2 \cdot 8^4 = 8^{12}. \tag{72}$$

According to Rule 1, the product of $a^x$ and $a^y$ will therefore be $a^{x+y}$. If $y$ is equal to zero, this becomes $a^{x+0}$, or,

$$a^x \cdot a^0 = a^x. \tag{73}$$

Solving this expression for $a^0$, Eq. (73) becomes

$$a^0 = \frac{a^x}{a^x}, \tag{74}$$

from which

$$a^0 = 1. \tag{75}$$

From the above is derived the second rule governing exponents:

*Rule* 2. *Any number (other than zero) with a zero exponent is equal to* 1.

If Rule 1, governing the multiplication of numbers having exponents,

is to hold true for all cases, it must also apply to numbers having negative exponents. Thus, according to Rule 1,

$$a^x \cdot a^{-y} = a^{x-y}. \tag{76}$$

For the special case where the exponents are numerically equal,

$$a^x \cdot a^{-x} = a^{x-x} = a^0. \tag{77}$$

However, according to Rule 2,

$$a^0 = 1. \tag{78}$$

Substituting in Eq. (77), the equivalent value of $a^0$ obtained from Eq. (78),

$$a^x \cdot a^{-x} = 1. \tag{79}$$

Dividing both members of Eq. (79) by $a^x$ (solving for $a^{-x}$),

$$a^{-x} = \frac{1}{a^x}. \tag{80}$$

From the above derivation, Rule 3 has been formulated.

*Rule 3. Any number with a negative exponent is equal to the reciprocal of the same number with a numerically equal positive exponent.*

According to this rule,

$$a^{-x} = \frac{1}{a^x}, \tag{80}$$

and similarly,

$$b^{-y} = \frac{1}{b^y}. \tag{81}$$

It follows, therefore, that

$$\frac{a^{-x}}{b^{-y}} = \frac{\frac{1}{a^x}}{\frac{1}{b^y}} = \frac{1}{a^x} \cdot \frac{b^y}{1} = \frac{b^y}{a^x}. \tag{82}$$

From the above, the following rule is derived.

*Rule 4. Any number having an exponent may be transferred from the numerator to the denominator of a fraction, or vice versa, by changing the sign of the exponent.*

Applying Rule 4, it is apparent that the fraction $\frac{10^5}{10^2}$ can be converted to simply $10^3$. Generalizing,

$$\frac{a^x}{a^y} = a^{x-y}, \tag{83}$$

from which Rule 5 is derived.

*Rule 5. Like numbers having exponents may be divided by subtracting exponents and affixing the resulting exponent to the common number.*

Thus,

$$\frac{10^{9}}{10^{-3}} = 10^{12}, \tag{84}$$

and

$$\frac{10^{-6}}{10^{4}} = 10^{-10}. \tag{85}$$

The process of raising a number to a power is called **involution.** Involution of numbers having exponents is accomplished by multiplying the exponent by the power to which it is to be raised. Thus,

$$(10^{4})^{2} = 10^{8}, \tag{86}$$

$$(10^{5})^{2} = 10^{10}, \tag{87}$$

$$(10^{-3})^{4} = 10^{-12}. \tag{89}$$

The process of extracting a root of a number is called **evolution.** Evolution of numbers having exponents is accomplished by reversing the above procedure, that is, the exponent is divided by the root which it is desired to extract. Thus,

$$\sqrt{10^{6}} = 10^{3}, \tag{90}$$

$$\sqrt[3]{10^{6}} = 10^{2}, \tag{91}$$

$$\sqrt[4]{10^{8}} = 10^{2} \tag{92}$$

The application of the operation of evolution to numbers having exponents has led to the use of *fractional* exponents. Fractional exponents are simply more practical methods of expressing the extraction of a root of a number that is to be raised to a power, that is, a number having an exponent. For example, Eq. (92) can be expressed as $10^{8/4}$. By performing the division indicated by the fraction, the term becomes $10^{2}$.

In converting numbers to numbers that are in exponent form, the following principles will serve to guide the student.

A whole number is converted to a number in exponent form by taking that part of the number up to and including the last digit and multiplying it by 10 raised to a power that is numerically equal to the number of ciphers between the last digit and the decimal point. Thus,

$$32{,}800{,}000 = 328 \cdot 10^{5}, \tag{92}$$

$$3{,}000 = 3 \cdot 10^{3}. \tag{93}$$

A decimal fraction is converted to exponent form by multiplying the digit portion of the expression by 10 raised to a *negative* exponent, which is equal numerically to the number of ciphers and digits from the decimal point up to, and including, the *last digit*. Thus,

$$0.0005 = 5 \cdot 10^{-4}, \tag{94}$$

$$0.00079800 = 798 \cdot 10^{-6}. \tag{95}$$

When the limits of accuracy permit, whole numbers are often taken to the nearest million or hundred thousand to facilitate conversion to exponential form. For example, 927,235,628 can be roughly taken as 927,000,000, or $927 \cdot 10^6$, with usually negligible sacrifice of accuracy. Similarly, 0.000893264 can be taken as 0.0009 or $9 \cdot 10^{-4}$.

One of the most common applications of exponents in radio work is in rapid conversion of units having metric prefixes (see Chap. 1) to units having other prefixes or no prefix at all. Since the prefixes differ according to the metric system, a conversion is accomplished by either multiplying or dividing by 10 or some power of 10. By the use of exponents, such conversion is reduced to the simple operation of multiplying the term in question by 10 raised to the proper positive or negative exponential power. Thus, to convert 6.2835 kilovolts to volts, simply multiply by $10^3$, or

$$6.2835 \text{ kv} = 6.2835 \cdot 10^3 \text{ v}. \tag{96}$$

Similarly, 1,003 $\mu$f becomes $1{,}003 \cdot 10^{-6}$ f, and so on. It is good practice in the solution of technical problems to eliminate decimal points altogether so far as possible by converting the terms to whole numbers multiplied by 10 to the proper power. The common mistake of misplacing the decimal point throughout the many operations of a complex problem is thus avoided. It is then necessary to count off the decimal places only once in the answer. Thus, 6.2835 kv in Eq. (96) can be converted to exponential form by multiplying by 10,000 or $10^4$, giving the whole number 62,835. However, a given value cannot casually be multiplied by whatever factor is convenient. In order to maintain the original value of the term, the resulting whole number must be *divided* by an amount equal to that by which it has been multiplied. This is indicated in exponential form as follows:

$$6.2835 \text{ kv} = 62{,}835 \cdot 10^{-4} \text{ kv}, \tag{97}$$

and

$$6.2835 \text{ kv} = 62{,}835 \cdot 10^{-1} \text{ v}. \tag{98}$$

Similarly,

$$0.003\ \mu\text{f} = 3 \cdot 10^{-3}\ \mu\text{f}, \tag{99}$$

and

$$0.003\ \mu\text{f} = 3 \cdot 10^{-9} \text{ f}. \tag{100}$$

The solution of a typical radio problem with the aid of exponents follows.

**Problem.** Given

$$X_c = 1/(2\pi f C), \tag{101}$$

where $X_c$ is in ohms, $2\pi$ is the constant 6.28, $f$ is 10 kc, $C$ is 0.004 $\mu$f. Convert all terms to basic units (cycles, farads), and solve for $X_c$.

**Solution.**

$$X_c = \frac{1}{6.28 \cdot 10{,}000 \cdot 0.000000004}. \tag{102}$$

Converting to exponents,

$$X_c = \frac{1}{628 \cdot 10^{-2} \cdot 10^{4} \cdot 4 \cdot 10^{-9}}. \tag{103}$$

Combining terms,

$$X_c = \frac{1}{628 \cdot 4 \cdot 10^{-7}}. \tag{104}$$

Moving exponent to numerator,

$$X_c = \frac{10^7}{628 \cdot 4}. \tag{105}$$

By arithmetic cancellation,

$$X_c = \frac{10{,}000{,}000}{628 \cdot 4} = \frac{2{,}500{,}000}{628}, \tag{106}$$

$$X_c = 3{,}980 \text{ ohms}. \tag{107}$$

The only arithmetical operations necessary for the solution of the above problems were the simple cancellations of step (106) and the small division of step (107). The extremely awkward multiplication and division of step (102) along with the attendant greater risk of error was completely avoided by the use of exponents.

**Logarithms.** In the history of mathematics, it was not long before the advantage of performing arithmetical calculations by means of exponents was recognized. Early in the seventeenth century, the limitations of exponential calculation led to the development of a more advanced system of computation known as "logarithms." Logarithms were invented by the Scotsman Lord Napier and are an outgrowth of the original theory of exponents. Logarithms are designed to simplify long computations by representing all real positive numbers as powers of some particular number. The exponents of these powers are called **logarithms** and are arranged in tables for convenient reference. In accordance with the principles of exponents, multiplication, division, involution and evolution of lengthy numbers are replaced by addition, subtraction, multiplication, and division, respectively, of the logarithms of the numbers.

The particular number that is used as the *base* throughout the system of logarithms was the subject of considerable discussion between Lord Napier and Henry Briggs. For various reasons, it was at first thought that logarithms to the base of an irrational number whose approximate value is 2.71828 would be an advantage. The system utilizing this base is called the **Napierian,** or **natural,** system of logarithms. The manifold advantages of logarithms to the base 10 were soon recognized, however.

The development of this system is due largely to Briggs, and it is in his honor that the system to the base 10 is called the **Briggs, or common,** system of logarithms. The latter system is more extensively used throughout the world and is used almost exclusively in radio engineering. All logarithmic work in this book will be according to the Briggs system.

The exponent of the power to which the number 10 must be raised in order to produce a given number is called the **logarithm of the given number.** Thus, the logarithm of 100 is 2, since $100 = 10^2$, or 10 raised to the second power.

A table of logarithms is simply a table of powers (exponents) to which the number 10 must be raised to obtain any number. A typical logarithm table computed to four decimal places is included in the Appendix. It is obvious that since this table includes a great many numbers that are not whole powers of the number 10 it is composed almost entirely of fractional decimal numbers. The logarithm of a number other than an integral power of 10 is composed of an integer and a fractional, or decimal part. The integral part of a logarithm is called the **characteristic,** and the decimal part is called the **mantissa.**

The characteristic of the logarithm of a number is obtained by observation of the number and is therefore not included in the table of logarithms. A table of logarithms is actually a table of the mantissas of logarithms.

The following rules serve to guide the student in determining the characteristic of the logarithm of a number:

*Rule 1. The characteristic of the logarithm of a number greater than 1 is either positive or zero and is 1 less than the number of digits in the integral part of the number.*

*Rule 2. The characteristic of the logarithm of a decimal is negative and is numerically 1 greater than the number of ciphers immediately following the decimal point.*

Logarithm is customarily abbreviated "log," and the abbreviation is used without a period. According to the above rules,

$$\text{characteristic of } \log 7590 = 3, \tag{108}$$
$$\text{characteristic of } \log 759.0 = 2, \tag{109}$$
$$\text{characteristic of } \log 75.90 = 1, \tag{110}$$
$$\text{characteristic of } \log 7.590 = 0, \tag{111}$$
$$\text{characteristic of } \log 0.7590 = -1, \tag{112}$$
$$\text{characteristic of } \log 0.0759 = -2, \tag{113}$$
$$\text{characteristic of } \log 0.00759 = -3, \tag{114}$$

and so on.

It should be noted that for decimals, such as those in (112), (113), and (114), *only the characteristic* of the logarithm is negative. The mantissa as taken from a log table remains positive, regardless of the sign of the

characteristic. Placing a negative sign before the entire logarithm of a decimal is therefore incorrect. This difficulty is overcome by adding 10 to the characteristic of the logarithm when the characteristic is negative. The characteristic is thereby made positive and agrees in sign with the mantissa. The subtraction of the number 10 is then indicated immediately following the logarithm in order to preserve the numerical value of the logarithm. For example, the characteristic of the logarithm of the decimal 0.4580 is $-1$. The mantissa as obtained from a log table is .6609. Therefore $\log 0.4580 = -1 + .6609$. This is written $9.6609 - 10$.

Often in handling extremely small decimal numbers a characteristic greater than $-10$ is encountered. In such cases, some multiple of 10 (20, 30, 40, or the like) is added to the characteristic and a like amount subtracted from the entire logarithm. It should be noted that it is not necessary to *perform* the actual subtraction. It suffices merely to *indicate* the subtraction. The necessity of performing the operation is obviated because the logarithm is eventually converted back to a number again.

The logarithms of numbers expressed by the same figures in the same order differ only in their *characteristics*. The *mantissas* of the logarithms of such numbers are the same. Thus

$$\log 8430. = 3.9258 \tag{115}$$

$$\log 843.0 = 2.9258, \tag{116}$$

$$\log 84.30 = 1.9258, \tag{117}$$

$$\log 8.430 = 0.9258, \tag{118}$$

$$\log 0.8430 = 9.9258 - 10, \tag{119}$$

$$\log 0.0843 = 8.9258 - 10. \tag{120}$$

The table of logarithms in the Appendix gives the mantissa to four decimal places of the common logarithms of all numbers from 1 to 1,000. Although log tables used in engineering work are usually computed to six or more decimal places and cover ranges from 1 to 10,000 or 100,000, the Appendix table will be found sufficiently accurate for the problems encountered in this text. The manner of obtaining the logarithms of numbers outside the range of 1 to 1,000 is discussed in a later section of this chapter.

The procedure of finding the logarithm of a number is best explained by taking a typical number and outlining the process step by step. Thus, to find the log of 458, the log table is used as follows. The letter $N$ in the Appendix log table is used to designate the vertical column of numbers from 10 to 99 inclusive. It also designates the horizontal row of numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The first two integers of a number are found in the vertical column under $N$. The third figure of the number is then located in the horizontal row opposite $N$. Thus, the number 45 is located in the vertical $N$ column and the number 8 in the horizontal row. The vertical column under 8 is then traced downward until it intersects with

the horizontal row opposite 45. The figure at this intersection is .6609, which is the mantissa of the desired logarithm. According to the rules governing characteristics (page 33), the characteristic of the log of 458 is 2. The logarithm of 458 is therefore 2.6609.

It will be noted that the table contains the mantissas not only of the logs of numbers expressed by three figures but also of logarithms expressed by four figures when the last figure is 0. Thus the mantissa of the log of 4,580 is .6609. The mantissa of the logarithm of 4,590 is found from the table to be .6618 It is apparent that the mantissa of the logarithm of a number *between* 4,580 and 4,590, such as 4,586, must lie between .6609 and .6618. The difference between the mantissas .6618 and .6609 is 9 ten-thousandths. This is called the **tabular difference.** Since the numbers 4,580 and 4,590 differ by 10, it may be assumed that each increase of 1 unit in progressing from 4,580 to 4,590 produces a corresponding increase of 0.1 of 9 ten-thousandths in the mantissa of the logarithm. This is not strictly true, but it is sufficiently accurate for most practical work. Consequently, 4,586 (6 added to 4,580) will add 0.6 of 9 ten-thousandths to the mantissa of the logarithm of 4,580. Six tenths of 9 ten-thousandths is approximately 5 ten thousandths. The mantissa of the logarithm of 4,586 is therefore .6609 + .0005 = .6614.

The process described above is called **interpolation.** Although finding the logarithm of a number may at first seem to be a cumbersome process, the student will find that he soon acquires speed. After a little practice, all necessary interpolations can be performed mentally, and the use of logarithms becomes an almost mechanical process.

**Antilogarithm** is the name given to the number which corresponds to a given logarithm. Thus, in the above illustration, it was found that the logarithm of 4,586 is 3.6614; thus, 4,586 is called the antilogarithm of 3.6614. Finding the antilogarithm when the logarithm is given is exactly the reverse of finding the logarithm when the number is given.

The first step in finding the antilogarithm of a given logarithm is to find the mantissas in the table next larger and next smaller than the mantissa of the given logarithm. Thus, if the given logarithm is 1.8529, the adjacent mantissas are found from the table to be .8525 and .8531. The numbers corresponding to these mantissas are 712 and 713, or 7,120 and 7,130. The tabular difference is .8531 − .8525, or .0006. Since the numbers 7,120 and 7,130 differ by 10, and the mantissas of their logarithms differ by 6 ten-thousandths, it is assumed sufficiently accurate to say that each increase of 0.0001 in the mantissa is produced by an increase of one sixth of 10 in the number. Thus, .8529 (the mantissa of the given logarithm) represents an increase of 0.0004 (4 ten-thousandths) over the mantissa (.8525) of the logarithm of the number 7,120 This corresponds to an increase of $\frac{4}{6}$ of 10, or approximately 6, in the number. Hence, the number corresponding to the mantissa .8529 is 7,126. The

number whose logarithm is the given logarithm 1.8529 is therefore 71.26. Since the given characteristic is 1, there are two integers to the left of the decimal point, in accordance with the rules governing characteristics.

Since the logarithms are exponents of the powers to which a common number (10) is to be raised, they obey all the rules that are applicable to exponents. Consequently, the processes of multiplication, division, involution, and evolution of cumbersome numbers can be accomplished by the addition, subtraction, multiplication and division of the logarithms of the numbers.

**Problem.** Multiply 0.0381, 0.00469, 0.981, and 5,420 by means of logarithms.

**Solution.** From the tables,

$$\log 0.0381 = 8.5809 - 10, \tag{121}$$
$$\log 0.00468 = 7.6702 - 10 \tag{122}$$
$$\log 0.981 = 9.9917 - 10, \tag{123}$$
$$\log 5420 = 3.7340 \tag{124}$$

Adding logs,

$$\text{sum of logs} = 29.9768 - 30 \tag{125}$$

or

$$\text{sum of logs} = 9.9768 - 10 \tag{126}$$

Then,

$$\text{antilog } 9.9768 - 10 = 0.948 \tag{127}$$

Therefore, the product of 0.0381, 0.00468, 0.981, and 5,420 is 0.948.

**Problem.** Divide 0.0000000507 by 0.00000785.

**Solution.** From the tables,

$$\log 0.0000000507 = 12.7050 - 20 \tag{128}$$
$$\log 0.00000785 = 4.8948 - 10, \tag{129}$$

Subtracting logs,

$$\text{difference} = 7.8102 - 10 \tag{130}$$

Then,

$$\text{antilog } 7.8102 - 10 = 0.00646. \tag{131}$$

Therefore,

$$0.0000000507 \div 0.00000785 = 0.00646 \tag{132}$$

In order to avoid subtracting a larger number from a smaller one as in the above problem the positive part of the logarithm of the dividend (128) is made to exceed that of the divisor (129). This is accomplished by adding 10 − 10, 20 − 20, and so on as the occasion demands.

**Problem.** Find the value of $0.0384^6$.

**Solution.**

$$\log 0.0384^6 = 6 \log 0.0384 \tag{133}$$

From the tables,

$$\log 0.0384 = 8.5843 - 10 \tag{134}$$

Then,

$$6 \log 0.0384 = 51.5058 - 60, \tag{135}$$

or,

$$6 \log 0.0384 = 1.5058 - 10. \tag{136}$$

Then,

$$\text{antilog } 1.5058 - 10 = 0.0000000003206. \tag{137}$$

Therefore,

$$0.0384^6 = 0.0000000003206. \tag{138}$$

**Problem.** Find the cube root of 0.1296

**Solution.**

$$\log 0.1296^{1/3} = \tfrac{1}{3} \log 0.1296. \tag{139}$$

From the tables,

$$\log 0.1296 = 9.1126 - 10. \tag{140}$$

Obviously, this logarithm is not evenly divisible by 3. Therefore a multiple of 10 is added to make the negative part of the logarithm a number evenly divisible by 3. Then step (140) becomes

$$\log 0.1296 = 29.1126 - 30. \tag{141}$$

Dividing by 3,

$$\frac{29.1126 - 30}{3} = 9.7042 - 10. \tag{142}$$

Then,

$$\text{antilog } 9.7042 - 10 = 0.506. \tag{143}$$

Therefore,

$$\sqrt[3]{0.1296} = 0.506. \tag{144}$$

**The Quadratic Equation.** An equation which, when simplified, contains the *square* of the unknown number, but no higher power, is called an **equation of the second degree,** or a **quadratic equation.** Since the unknown is present raised to the second degree, it follows that the solution of such an equation involves extracting the square root of the remaining terms.

Thus,

$$8x^2 = 99 - 3x^2. \tag{145}$$

Transposing,

$$11x^2 = 99, \tag{146}$$

and

$$x^2 = 9. \tag{147}$$

Extracting the square root of both sides,

$$x = 3. \tag{148}$$

However, the solution of step (148) could also be $x = -3$. This follows, since $(+3)(+3) = 9$, and $(-3)(-3) = 9$. Step (148) therefore, becomes

$$x = \pm 3. \tag{149}$$

An equation of the general form of Eq (145) is called a **pure quadratic equation.** A pure quadratic equation may be defined as one that contains only the second power of the unknown

*Every pure quadratic equation has two roots, numerically equal, but opposite in sign.*

A quadratic equation that contains the unknown in the first power as well as in the second power is called an **affected quadratic equation.** Thus,

$$x^2 \quad 2x - 10 \tag{150}$$

is an affected quadratic equation, since the unknown $x$ is present in both the first and second powers. Affected quadratic equations can always be reduced to the general form of

$$ax^2 + bx + c = 0, \tag{151}$$

where the coefficients $a$ and $b$ and the *absolute term* $c$ may represent any number whatsoever.

Quadratic equations may be solved by any of a number of different methods, including factoring, completing the square, and the use of a formula Each of the methods has its advantages If the factors of an equation are easily found, the method of solution by factoring is to be preferred. In many cases however, factoring of a complex term becomes rather cumbersome and difficult. In such cases, the solution is often best obtained by means of the method of completing the square. This method, however, is also not always easy to apply The one method that is applicable in *all* cases is the method of solution by formula This is, therefore, the method that will be discussed here.

The general quadratic of Eq. (151) represents *any* quadratic equation whatever containing only one unknown. The solution of Eq (151) can be made the solution of *any* quadratic equation simply by substituting for the coefficients $a$ and $b$, and the absolute term $c$ in the result obtained when Eq. (151) is solved for $x$. Equation (151) has been solved by the method of completing the square with the result,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{152}$$

The general procedure for the solution of quadratic equations is first to reduce the equation to the general form of Eq. (151). The numerical equivalents of the coefficients $a$ and $b$ and the absolute term $c$ are then substituted in Formula (152). Since the radical sign is prefixed by both a plus and minus sign, the result is two roots, either one of which satisfies the conditions of the equation

**Problem.** Solve

$$2x^2 + 5x = -2. \tag{153}$$

**Solution.** Converting to the general form of Eq. (151) by transposition,

$$2x^2 + 5x + 2 = 0. \tag{154}$$

Referring to the general form of Eq. (151), 2 is the numerical equivalent of $a$ (coefficient of $x^2$), 5 is the numerical equivalent of $b$ (coefficient of $x$); and 2 is the numerical equivalent of $c$ (the absolute term). Substituting in Eq. (152),

$$x = \frac{-5 \pm \sqrt{25 - 4(2)(2)}}{2(2)}. \tag{155}$$

Then,

$$x = \frac{-5 \pm \sqrt{25 - 16}}{4}, \tag{156}$$

$$x = \frac{-5 \pm \sqrt{9}}{4} = \frac{-5 \pm 3}{4}. \tag{157}$$

Therefore,

$$x = \frac{-5 + 3}{4} = -\frac{2}{4} = -\frac{1}{2} \tag{158}$$

and

$$x = \frac{-5 - 3}{4} = -\frac{8}{4} = -2. \tag{159}$$

The roots of Eq. (153) are therefore $-2$ and $-\frac{1}{2}$.

**Vector Algebra.** One of the most useful forms of mathematical representation in radio work is the graph. By means of a graph, it is possible to visualize the relation of abstract parameters of an electrical circuit which cannot be conceived in any other way.

Figure 9 consists of two lines $XX'$ and $YY'$ at right angles to each other. Each of these lines is subdivided into unit divisions, forming two scales. The point of intersection of the two lines is designated as zero. All subdivisions on the $YY'$ scale *above* the $XX'$ scale are positive units; and all those *below* the $XX'$ scale are negative. Similarly, all subdivisions on the $XX'$ scale *to the right* of the $YY'$ scale are positive units; and all those *to the left* of the $YY'$ scale are negative units. Any point $P$ in the plane containing the lines $XX'$ and $YY'$ can then readily be located in this graph by dropping perpendiculars from the point to the $XX'$ scale and the $YY'$ scale. Thus, point $P$ in Fig. 9 is located by dropping the perpendicular $PA$ to $YY'$ and the perpendicular $PB$ to the $XX'$ scale.

The perpendicular distance from $P$ to $YY'$, represented by line $PA$ in Fig. 9, is called the **horizontal coordinate** of point $P$. When point $P$ is to the right of the $YY'$ scale, the horizontal coordinate is positive; when it is to the left of the $YY'$ scale, the horizontal coordinate is negative. Similarly, when point $P$ is above the $XX'$ scale, the **vertical coordinate** is positive; when $P$ is below the $XX'$ scale, the vertical coordinate is negative.

Each of the scales $XX'$ and $YY'$ is called an **axis** of the coordinate system. The intersection of the axis is called the **origin** of the system.

The horizontal coordinate of a point is its **abscissa**; the vertical coordinate of a point is its **ordinate.** Collectively, the abscissa and ordinate of any point are termed the **rectangular coordinates** of that point. To **plot a point** means to lay off its abscissa and ordinate on a coordinate system and locate the point with a dot or cross mark.

In any discussion of a given problem, a **constant** is a term whose numerical value does not change throughout the discussion. A **variable** is a term that may assume different values throughout the discussion. If an equation contains two variables, the variables are **related variables,**

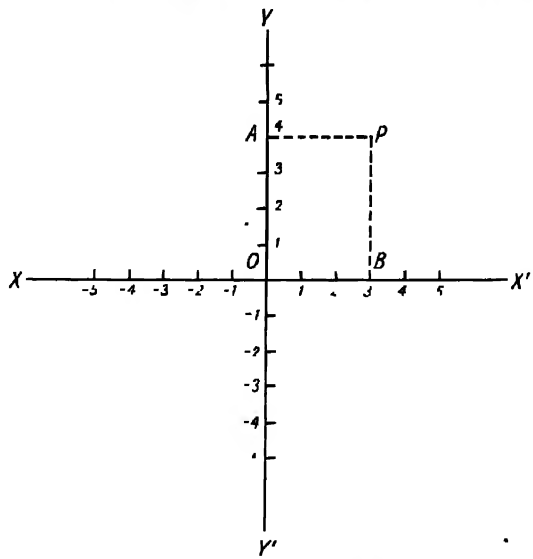



FIG. 9. A coordinate system. $PA$ and $PB$ are the rectangular coordinates of point $P$. $PA$ is the horizontal coordinate, or abscissa. $PB$ is the vertical coordinate, or ordinate.

since the value of one variable at any instant will depend upon the value of the other variable at that instant. Thus, if the value of the first variable is known, the value of the remaining variable can be found by solving the equation. The first variable is the **independent variable,** and the second variable is the **dependent variable.**

The equation

$$E = IR \tag{160}$$

is the mathematical relationship between voltage $E$, current $I$, and resistance $R$ in a simple electrical circuit. It is known as Ohm's law and is discussed in detail in Chap. IV. If $R$ is a constant, any change in $E$ will cause a change in $I$. If a numerical value of 10 is assigned to $R$ and $E$ is changed from 0 to 10 v in steps of 1 v, the current $I$ for each different voltage can readily be computed by solving Eq. (160) for $I$ for each

different value of voltage $E$. In such a problem, $E$ is the independent variable, and $I$, the dependent variable. Thus, when $E = 0$,

$$I = \frac{E}{R} = \frac{0}{10} = 0. \tag{161}$$

When $E = 1$,

$$I = \frac{E}{R} = \frac{1}{10} = 0.1. \tag{162}$$

When $E = 2$,

$$I = \frac{E}{R} = \frac{2}{10} = 0.2. \tag{163}$$

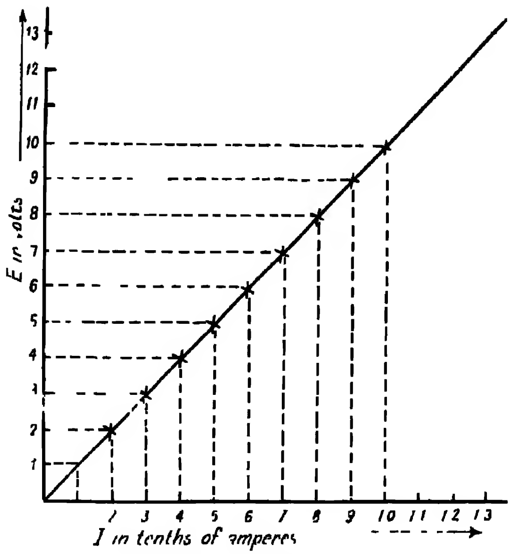



FIG. 10.—Graph of the linear equation $E = IR$ when $R$ is a constant. From this graph, the value of $I$ can be found for any value of $E$ within the limits of the graph.

In similar fashion, the corresponding values of $I$ when $E = 3$, 4, 5, 6, 7, 8, 9, and 10 v are found to be 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0, respectively. $I$ in this equation is therefore said to be a *function* of $E$, since the value of $I$ at any instant depends upon the value of $E$ at that instant. The above values of $E$ and $I$ can be taken as the rectangular coordinates of a coordinate system. The various solutions of Eq. (160) can then be plotted in a graph. This is shown in Fig. 10. The values of $E$ are plotted on the $YY'$ axis. In order to make a more easily readable graph, the values of $E$ are scaled in units (volts), and the values of $I$ are scaled in tenths of units (tenths of amperes). Since Eq. (160) is a *simple* equation, its graph is found to be a straight line. The graphs of all simple equations are straight lines. For this reason, they are often called **linear equations.**

The advantage of a graph, such as that of Fig. 10, is that for this particular circuit the current $I$ that results from *any* voltage $E$ can readily be found directly from the graph. Thus, if $E$ is 2.5 v, the abscissa intersecting the $E$ scale is traced out until the point of intersection with the graph is found. From this point, the ordinate is traced vertically downward until it intersects the $I$ scale. The point of intersection will be found to be 0.25 amp.

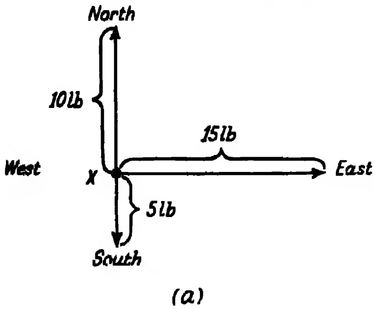



Fig. 11(a) Forces in three different directions are being exerted on the ball $x$.

Another interesting application of graphs is in the solution of forces about a given point. If it is assumed that point $X$ in Fig. 11(a) is a ball that is free to move in any direction, the actual direction in which the ball will move when a number of different forces are applied to it from different directions can be calculated graphically. Thus, ball $x$ in Fig. 11(a) is subjected to three different forces. A force of 10 lb is tending to

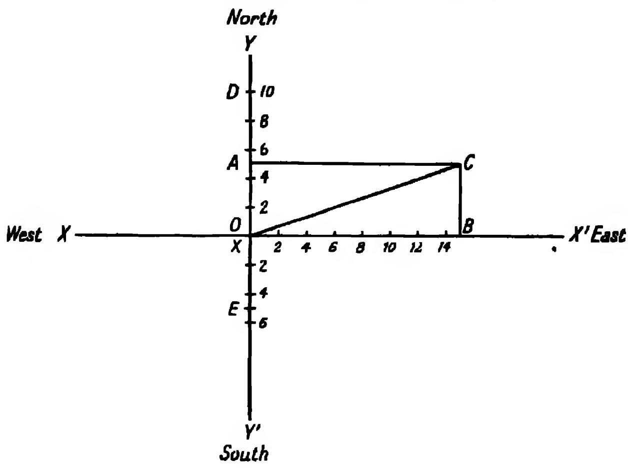



Fig. 11(b). The forces of Fig. 11(a) graphed on a coordinate system.

move it to the north; at the same time, a force of 5 lb is tending to move it southward; simultaneously, a force of 15 lb is tending to move the ball eastward. In which direction will the ball move?

The first step in the solution of such a problem consists of laying off the forces involved on the $XX'$ and $YY'$ axes of a graph. This is shown in Fig. 11(b). The eastward force of 15 lb is scaled off on the $XX'$ axis. In order to show the magnitude of the forces, both axes of the coordinate

system are calibrated in pounds. Thus, a force of 15 lb to the eastward is represented by a line of the $XX'$ axis equal in length to 15 divisions on the $XX'$ scale. Similarly, the force to the northward is shown as a line on the $YY'$ scale 10 divisions in length. Since the northward force is 90° removed from the eastward force, it falls on the $YY'$ axis, which is at right angles to the $XX'$ axis. The southward force is 90° removed from the eastward force in a direction opposite to that of the northward force, or 180° from the northward force. It therefore also falls on the $YY'$ axis. Since this force is opposite in direction from the northward force, it is shown as a *negative* quantity, that is, *below* the $XX'$ scale. This is shown in Fig. 11(b) as a line on the $YY'$ scale 5 divisions in length, counting from zero downward.

Since the northward and southward forces are opposite to each other, they can be added algebraically to obtain the total force acting along the $YY'$ axis. The result is 5 lb acting to the northward and is indicated by the line $OA$ in Fig. 11(b). The three forces acting upon the ball have now been reduced to two forces acting at right angles to each other. As will be shown later in this chapter, any number of forces acting in *any* direction upon the same point can be resolved by geometric or trigonometric means to two major forces acting upon the point at right angles to each other.

One of the principal laws of mechanics states that:

*The resultant of two forces acting upon a point may be represented by the diagonal of a parallelogram two of the adjacent sides of which represent the two forces.*

A parallelogram is therefore constructed on the two forces $OA$ and $OB$ in Fig. 11(b). The diagonal $OC$ then represents the total force acting upon ball $r$. The ball will move in the direction of $OC$ with a force equal to the length of line $OC$. This force can be taken directly from the graph utilizing scalar divisions equal to the divisions of the $XX'$ and $YY'$ scales.

Since the two component forces are at right angles to each other, the parallelogram is a rectangle. Triangle $OCB$ is therefore a right triangle. Consequently, the magnitude of force $OC$ can be calculated directly from the values of $OA$ and $OB$ by means of the geometric theorem known as the **Pythagorean theorem,** named after the Grecian mathematician (525 B.C.) who postulated it:

*The square of the hypotenuse of a right triangle is equal to the sum of the squares of the other two sides.*

The magnitude of force $OC$ can therefore be expressed

$$OC^2 = OB^2 + OA^2, \tag{164}$$

or

$$OC = \sqrt{OB^2 + OA^2}. \tag{165}$$

Substituting from Fig. 11(b) the equivalent values of $OB$ and $OA$,

$$\overline{OC} = \sqrt{15^2 + 5^2}, \tag{166}$$

$$\overline{OC} = \sqrt{225 + 25}, \tag{167}$$

$$\overline{OC} = \sqrt{250} = 15.8 \text{ lb.} \tag{168}$$

Equation (165) can be rearranged in terms of the original forces as follows:

$$\overline{OC} = \sqrt{OB^2 + (OD - OE)^2}, \tag{169}$$

where $OD$ and $OE$ are the original forces, as shown in Fig. 11(b).

The forces acting about a given point are known as **vectors.** Each of the individual forces acting upon the point is known as a **component vector,** and the net force acting upon the point as a result of the component forces is called the **resultant vector.** Thus, in the problem illustrated in Fig. 11(b), $OB$, $OD$, and $OE$ are the component vectors; $OC$ is the resultant vector.

If the direction and magnitude of each force is known, any number of forces acting upon a point can be resolved into two major forces acting upon the point at right angles to each other. These two forces can therefore be represented as acting along the $XX'$ and $YY'$ axes of a coordinate system. Equation (169) can accordingly be generalized to apply to all cases. Stated as a rule:

*The resultant vector of the forces acting upon a given point is equal to the square root of the sum of the algebraic sum of the vectors acting along the $YY'$ axis squared and the algebraic sum of the vectors acting along the $XX'$ axis squared.*

Expressed mathematically,

$$X = \sqrt{A^2 + B^2}, \tag{170}$$

where $X$ = resultant vector;
$A$ = algebraic sum of vectors acting along $XX'$ axis;
$B$ = algebraic sum of vectors acting along $YY'$ axis.

Resolving a number of forces acting in various directions into two major forces acting at right angles to each other involves the use of trigonometry and is discussed in the following section.

## *TRIGONOMETRY*

The word "trigonometry" is derived from the Greek words *trigon*, "a triangle," and *metrein*, "to measure." The science of trigonometry is concerned chiefly with the relation of certain lines in a triangle. It forms the basis of the mensuration used in all branches of engineering.

**Trigonometric Functions.** In Fig. 12, if the line $AD$ is moved about the point $A$ as indicated by the arrow, it generates the angle $A$ (angle $EAD$). If from the line $AD$, at various points $B$, $B'$, $B''$, a series of perpendiculars $BC$, $B'C'$, $B''C''$ are let fall to line $AE$, a series of similar triangles $ABC$, $AB'C$, and $AB''C''$ are formed. Since these triangles are similar, their corresponding sides are proportional. That is,

$$\frac{BC}{AB} = \frac{B'C'}{AB'} = \frac{B''C''}{AB''} \tag{171}$$

$$\frac{BC}{AC} = \frac{B'C'}{AC'} = \frac{B''C''}{AC''} \tag{172}$$

$$\frac{AB}{AC} = \frac{AB'}{AC'} = \frac{AB''}{AC''}. \tag{173}$$

Similarly, any other ratio of the sides of each of these triangles is equal to the ratio of the same sides of the remaining triangles. Obviously, the lengths of the sides of these triangles is not the determining factor, since all the triangles are not of the same size. The ratios of the sides of these triangles remain unchanged as long as the angle $A$ remains unchanged. As the angle is made larger or smaller, the numerical value of the ratio of the sides of the triangles changes. The ratio of a given pair of sides in any one triangle, however, is *always* equal to the ratio of the corresponding pair of sides of any of the remaining triangles. The triangles in Fig. 12 were constructed by dropping *perpendiculars*. Hence, the triangles in question are *right* triangles. Each of the ratios in the triangles of Fig. 12 are therefore *functions* of the angle $A$. Each of the acute angles in a right triangle has six functions. The six functions of an angle are given specific names in trigonometry. Thus, the six functions of angle $A$ in Fig. 12 are as follows:

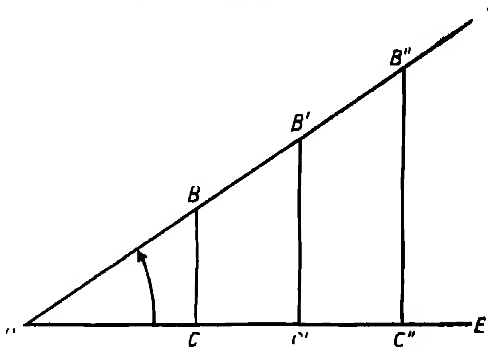


FIG. 12. As the line $AD$ moves in the direction indicated by the arrow, the angle $A$ ($= EAD$) is generated.

*sine* of $A$, abbreviated sin $A$;
*cosine* of $A$, abbreviated cos $A$;
*tangent* of $A$, abbreviated tan $A$,
*cotangent* of $A$, abbreviated cot $A$,
*secant* of $A$, abbreviated sec $A$;
*cosecant* of $A$, abbreviated csc $A$.

The trigonometric functions are defined as follows:

$$\sin A = \frac{\text{opposite side}}{\text{hypotenuse}} = \frac{a}{c},$$

$$\cos A = \frac{\text{adjacent side}}{\text{hypotenuse}} = \frac{b}{c},$$

$$\tan A = \frac{\text{opposite side}}{\text{adjacent side}} = \frac{a}{b},$$

$$\cot A = \frac{\text{adjacent side}}{\text{opposite side}} = \frac{b}{a},$$

$$\sec A = \frac{\text{hypotenuse}}{\text{adjacent side}} = \frac{c}{b},$$

$$\csc A = \frac{\text{hypotenuse}}{\text{opposite side}} = \frac{c}{a}.$$

The student should thoroughly memorize these definitions of the trigonometric functions, since they are the foundation upon which the entire study of trigonometry is built. The ratios of the small letters $a$, $b$, and $c$ in the above definitions refer to the sides of the triangle $ABC$ in Fig. 13. The functions defined are those of angle $A$ in this triangle.

The values of the trigonometric functions of any angle from 0 to 90° have been computed and tabulated. A typical engineering trigonometric table gives the values of the functions to six decimal places for every *minute* of arc from 0 to 90°. Values of functions for seconds of arc are found by interpolation in a manner similar to that used with logarithm tables. A table of trigonometric functions to four decimal places for every degree from 0 to 90° is included in the Appendix of this book. This table will suffice for problems where extreme accuracy is not required and can be used for the problems encountered throughout the text. Where greater accuracy is necessary, the student should refer to a set of practical trigonometric tables.

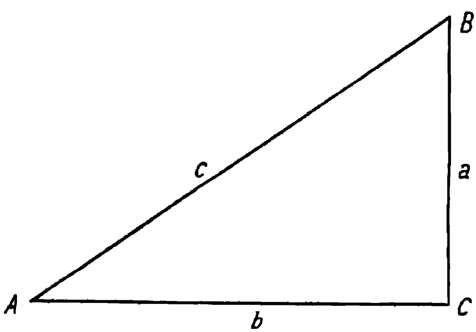


FIG. 13 The reader should refer to this triangle in defining the trigonometric functions of an angle

**Solution of Vectors.** Trigonometry is of tremendous practical value in the solution of vector diagrams of the type encountered in a-c theory,

as will be seen in Chap. IX. Addition of vector quantities is easily accomplished by trigonometric means in problems that would be insoluble by application of more elementary mathematics

The application of the trigonometric functions to the solution of a vector problem is best explained by the step-by-step solution of a typical problem. The best illustration of such a problem is again the case of a ball that is free to move in any direction. In Fig. 14, a ball is acted upon by four different forces, each tending to force the ball to move in a different direction. The first force $OA$ is exerting a pull of 10 lb to the northward. The second force $OB$ is exerting a pull of 20 lb in a direction 45° from north. The third force $OC$ is exerting a pull of 14 lb in a direction 120° from north. The fourth force $OD$ is exerting a pull of 5 lb in a direction 200° from north. The problem is to ascertain in which direction the ball will move and with what force as a result of these various forces exerted upon it.

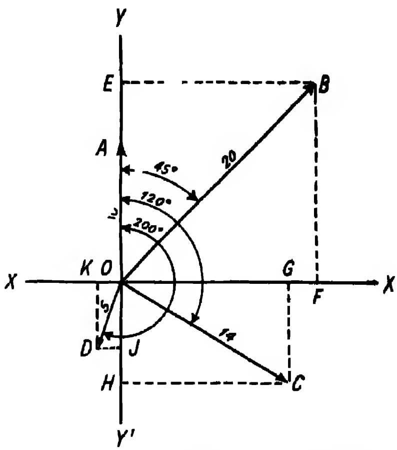


FIG. 14. A vector graph of the forces acting upon a point.

The solution of the problem resolves itself into several major steps. The ball is first graphically located in the center of a coordinate system. The $YY'$ axis is taken as lying in a north-south direction. Since the axes of a coordinate system are at right angles to each other, it follows, therefore, that the $XX'$ axis is in the east-west direction. Each of the applied forces is then shown on the graph in its proper relative position. The vectors are, of course, made the proper length consistent with the number of pounds of force that they represent. The second step in the solution is to resolve each of the vectors which does not fall on either of the coordinate axes into two component vectors which lie on the $XX'$ and $YY'$ axes. All the $XX'$ vectors and all the $YY'$ vectors are then separately added algebraically. The two major component vectors resulting from this addition are combined into a parallelogram of forces which is solved to find the resultant vector.

The graphical location of the original vectors is shown in Fig. 14. The three vectors $OB$, $OC$, and $OD$, which do not fall on either of the coordinate axes, are resolved into two component vectors by treating each as the diagonal of a parallelogram whose sides are the axes of the graph. Thus, a parallelogram is constructed about vector $OB$ by dropping perpendiculars from point $B$ to the $XX'$ axis and the $YY'$ axis. These

are shown in Fig. 14 by dotted lines *EB* and *FB*. Similarly, parallelograms *OGCH* and *OJDK* are constructed about vectors *OC* and *OD*, respectively. Vector *OB* can then be thought of as composed of component vectors *OE* and *OF*. Similarly, vector *OC* is composed of component vectors *OG* and *OH*, and vector *OD* is composed of component vectors *OJ* and *OK*.

It is at once apparent that the numerical value of each of the six component vectors cannot be obtained by ordinary geometric means. It is at this point that the trigonometric functions are used to advantage. In parallelogram *OEBF*, *OFB* is a right triangle. Angle *EOF* is 90°, since *EO* and *OF* lie on the coordinate axes. Since angle *EOB* is 45°, by subtraction angle *BOF* in triangle *OBF* is also 45°. Then,

$$\sin BOF = \frac{BF}{OB}. \tag{174}$$

Solving for *BF*

$$BF = OB \sin BOF \tag{175}$$

Substituting,

$$BF = 20 \sin 45° \tag{176}$$

From the trigonometric table in the Appendix

$$\sin 45° = .7071 \tag{177}$$

Hence,

$$BF = 20(.7071) = 14.142 \tag{178}$$

Similarly,

$$\cos BOF = \frac{OF}{OB} \tag{179}$$

and

$$OF = OB \cos BOF. \tag{180}$$

Substituting,

$$OF = 20 \cos 45° . \tag{181}$$

From table,

$$\cos 45° = .7071, \tag{182}$$

and

$$OF = 20(.7071) = 14.142. \tag{183}$$

Since *OEBF* is a parallelogram, *BF* = *EO*. Therefore,

$$EO = 14.142. \tag{184}$$

Angle *EOC* is given as 120°. Since angle *EOG* is 90°, it follows, by subtraction, that in triangle *OGC*, angle *GOC* is 30°. Then,

$$\sin GOC = \frac{GC}{OC}. \tag{185}$$

Then,

$$GC = OC \sin GOC. \tag{186}$$

Substituting,

$$GC = 14 \sin 30^\circ. \tag{187}$$

From the table,

$$\sin 30^\circ = .5, \tag{188}$$

and

$$GC = 14(.5) = 7. \tag{189}$$

But $GC = OH$, since $OGCH$ is a parallelogram. Therefore,

$$OH = 7. \tag{190}$$

Similarly,

$$\cos GOC = \frac{OG}{OC}. \tag{191}$$

Then,

$$OG = OC \cos GOC. \tag{192}$$

Substituting,

$$OG = 14 \cos 30^\circ. \tag{193}$$

From the table,

$$\cos GOC = .8660, \tag{194}$$

and

$$OG = 14(.8660) = 12.124. \tag{195}$$

Angle $EOD$ is given as $200^\circ$. Since angle $EOJ$ is $180^\circ$ (being a straight line), it follows by subtraction that in triangle $ODJ$, angle $DOJ$ is $20^\circ$. Then,

$$\sin DOJ = \frac{DJ}{OD}. \tag{196}$$

Solving,

$$DJ = OD \sin DOJ. \tag{197}$$

Substituting,

$$DJ = 5 \sin 20^\circ. \tag{198}$$

From the table,

$$\sin 20^\circ = .3420, \tag{199}$$

and

$$DJ = 5(.3420) = 1.71. \tag{200}$$

But $DJ = OK$, since $OJDK$ is a parallelogram. Therefore,

$$OK = 1.71. \tag{201}$$

In like manner,

$$\cos DOJ = \frac{OJ}{OD}. \tag{202}$$

Then,

$$OJ = OD \cos DOJ. \tag{203}$$

Substituting,

$$OJ = 5 \cos 20°. \tag{204}$$

From the table,

$$\cos 20° = .9397, \tag{205}$$

and

$$OJ = 5(.9397) = 4.6985. \tag{206}$$

All the vectors in the $XX'$ axis are now added algebraically. Since the vector $OK$ is to the left of the $YY'$ axis, it is diametrically opposed to the vectors to the right of the $YY'$ axis and must therefore be given a negative sign. Then,

| | | |
|---|---|---|
| vector $OF$ | 14.142 | (207) |
| vector $OG$ | 12.124 | (208) |
| vector $OK$ | 1.71 | (209) |
| Total $XX'$ vectors | 24.556. | (210) |

All the vectors of the $YY'$ axis are now added. The vectors below the $XX'$ axis, being opposite in direction to those above the $XX'$ axis, are given a negative sign. Then,

| | | |
|---|---|---|
| vector $OA$ | 10.0 | (211) |
| vector $EO$ | 14.142 | (212) |
| vector $OJ$ | 4.6985 | (213) |
| vector $OH$ | 7.0 | (214) |
| Total $YY'$ vectors | 12.4435. | (215) |

All the vectors concerned in the problem have now been resolved into two major vectors acting at right angles to each other. These vector forces are one of 12.4435 lb exerting a pull due north on the ball $x$ (since it is a positive vector quantity) and one of 24.556 lb exerting a pull due east on the ball. These two vectors are plotted on a separate coordinate system (Fig. 15) for clearness. They are shown as vector $OA$ 12.4435 and vector $OB = 24.556$ in Fig. 15.

A parallelogram of forces is now constructed having sides $OA$ and $OB$. The diagonal $OC$ is drawn. Triangle $OCB$ is a right triangle, since the parallelogram is a rectangle. Then,

$$\tan COB = \frac{CB}{OB}. \tag{216}$$

Substituting,

$$\tan COB = \frac{12.4435}{24.556}, \tag{217}$$

and

$$\tan COB = .5067. \tag{218}$$

From the table, the angle whose tangent is .5067 is approximately 27°. The resultant movement of the ball is therefore in a direction 27° north of east, or 63° from north.

By the Pythagorean theorem,

$$OC = \sqrt{OB^2 + OA^2}. \tag{219}$$

Substituting,

$$OC = \sqrt{(24.556)^2 + (12.4435)^2}. \tag{220}$$

and

$$\overline{OC} = \sqrt{754.01} = 27.4 \text{ lb.} \tag{221}$$

The ball will therefore move in a direction 63° from north with a force of 27.4 lb as a result of the various forces applied to it.

In radio, many problems require the use of vectors for their solution. Despite the fact that a physical distribution of forces does not always exist, it is often convenient to handle abstract quantitative electrical units in the same manner as physical forces. Thus, whenever the difference between two or more electrical quantities can be expressed in degrees as well as absolute value their relation can be represented graphically. Such quantities can then be combined vectorially to ascertain the net result in the electrical circuit of the original electrical forces. Whenever units are combined in this manner, they are said to have been added vectorially. Thus, in the problem above, the final resultant force is called the **vector sum** of the original forces. The application of vectors to electrical circuits is discussed in detail in Chap. IX.

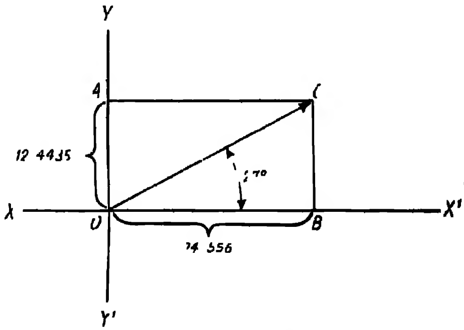



FIG. 15.—The resultant vector $OC$ is derived from the major component vectors $OA$ and $OB$.

*QUESTIONS AND PROBLEMS*

**1.** In the equation $E/E' = R/R'$, the values of $E$, $E'$ and $R'$ are known. Find $R$ in terms of the remaining values.

**2.** Given the equation $1/C = 1/C_1 + 1/C_2$, solve for $C_1$.

**3.** Given the simultaneous equations $EI = 800$ and $E = 25I^2$, find $E$ and $I$.

**4.** Given the formula $f = 1/(2\pi\sqrt{LC})$. When $f$, $\pi$, and $C$ are known, what is the value of $L$?

**5.** Given the formula

$$Z = \sqrt{R^2 + (X_L - X_C)^2},$$

solve for $X_L$.

**6.** The perimeter of a right triangle is 36 in One leg of the triangle is 3 in. longer than the other leg. How long is each side of the triangle?

**7.** Divide the product of 0.000000380 and 0.0000054 by 183,000,000 by the use of exponents.

**8.** By the use of logarithms, divide the product of 4,378 and 2,398 by the product of 992 and 6,042

**9.** Given the formula

$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}},$$

construct a graph showing the relationship of $R$ and $R_1$ if $R_2$ is equal to 5 and $R_1$ is increased from 1 to 10.

**10.** If $Z$, $X$, and $R$ are the hypotenuse and two legs of a right triangle, respectively, what is the angle formed by $Z$ and $R$ if $Z = 5$, $X = 3$, and $R = 4$?

## Chapter III

# CHEMICAL PRODUCTION OF ELECTRIC CURRENTS

Electricity can be generated in a variety of ways, including chemical means, mechanical means, the action of heat, friction, and pressure and expansion. Electricity is generated mechanically by application of the laws of electromagnetic induction, and such generators are discussed in Chap. VI. Chemical action and mechanical action are the only methods that are commercially practicable for producing electricity in sufficient quantities for power consumption. The remaining methods are of interest for applications other than power production or simply as laboratory phenomena.

The generation of electricity by heat is the basis of the **thermocouple unit.** It has been discovered that when pairs of certain dissimilar metals are brought into electrical contact and heat is applied, an **electromotive force** (abbreviated emf) is generated which is proportional to the temperature. This principle is utilized to advantage for the construction of delicate electrical measuring instruments (as in measuring temperature and current) but has little significance as a generator of electricity for power consumption purposes.

Static electricity can easily be generated by friction between two dissimilar substances and was first demonstrated by the ancient Greeks in their experiments with a piece of silk and a glass rod. This phenomenon is an interesting laboratory experiment but has no commercial importance.

The generation of electricity by pressure and expansion methods was first discovered in crystals of quartz, Rochelle salts, and certain other minerals. Such crystals were found to produce an emf when mechanical pressure was alternately applied to and removed from their opposing surfaces. Crystals find various advantageous applications in radio work. As a source of electrical power in other than minute quantities, however, they have no commercial value.

The chemical production of electricity is the result of immersing two different metals in an electrolyte. Since the electrolyte, being liquid, must be kept in a container, or cell, the name **electric cell** was derived for such a unit. When several such cells are connected in order to obtain a greater electrical output, the resulting combination is called a **battery** of cells. Usually, a battery of cells is series connected, as will be seen

later. Thus, the so-called automobile storage battery is actually a battery of cells connected in series.

There are two principal types of electric cells – the primary cell and the secondary cell. A **primary cell** is one that generates electricity by chemical action when a closed electric circuit is connected to its terminals. As a result of the chemical action, some of the components of the cell are consumed. Eventually, the components of a primary cell are completely used up, or exhausted, and the cell, as such, is useless. A **secondary cell** is one in which no chemical action can occur until after an electric current has been passed through the cell. Passing an electric current through a secondary cell is called **charging** the cell. Once the cell is charged, it can be connected to an external closed electric circuit, and electricity will then be generated by chemical action within the cell. When the cell is discharged in this manner, a point is eventually reached at which the cell components (electrodes and electrolyte) are no longer capable of chemical action. The cell must then be recharged, that is, an electric current must again be passed through it in the reverse direction. This cycle of charge and discharge can be repeated indefinitely without disintegration of the cell components. The cell, however, is subject to a certain amount of normal deterioration due to the harmful effects of extraneous chemical action, evaporation and so on.

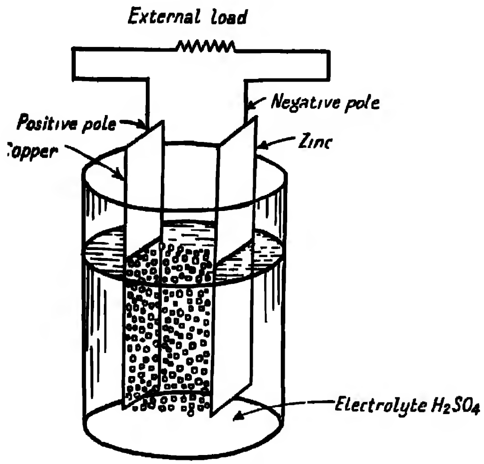



Fig. Elementary voltaic cell

## *THE PRIMARY CELL*

**Constructional Details.** The usual form of primary cell consists of a zinc and a copper plate immersed in a dilute solution of sulphuric acid. This type of cell is also often called a **voltaic** cell after the Italian physicist Volta who discovered it in 1800. An elementary primary cell is shown in Fig. 16. When the zinc and copper plates are connected externally by means of a conducting wire, a chemical action immediately takes place within the cell. The sulphuric acid in the electrolyte attacks the zinc, forming zinc sulphate. This action results in a predominance of

electrons in the zinc electrode, causing it to have a negative charge. A difference of potential consequently exists between the zinc and the copper electrodes, and electrons move through the wire from the zinc to the copper. During this process, the zinc plate disintegrates and eventually is entirely eaten away.

The electrolyte simultaneously is slowly decomposed, part of it joining with the zinc to form zinc sulphate and the remaining element hydrogen being released in the form of a gas. If the chemical action is slow (very little current flowing), the released hydrogen bubbles to the surface of the electrolyte and escapes into the air. More often, however, the hydrogen is attracted to the negative (with respect to the hydrogen ions) copper electrode. If the chemical action is fairly rapid, the hydrogen does not rise to the surface fast enough to keep the surface of the copper plate clear. The copper plate then is gradually covered with a film of hydrogen gas which increases the internal resistance of the cell. This process is called **polarization** and greatly limits the efficiency and life of the cell.

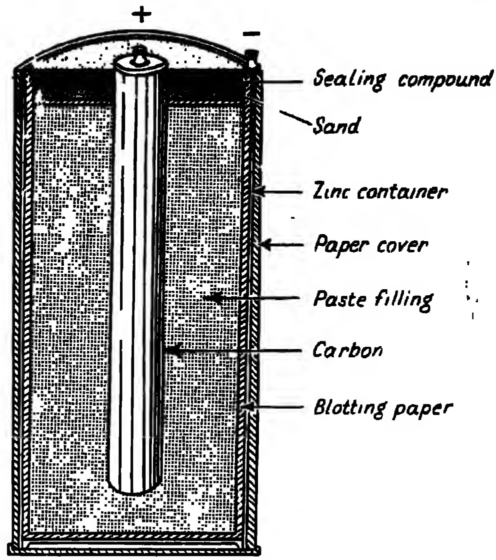



FIG. 17 Cross-sectional view of dry cell.

Polarization is usually counteracted by introducing oxygen in some form into the cell. Hydrogen readily combines with oxygen to form water and polarization is thus prevented. It was subsequently discovered that the life of a primary cell could be greatly prolonged by utilizing carbon instead of copper for the positive electrode. Because carbon is porous, it readily absorbs oxygen from the air, permitting it to combine with the hydrogen and thus delaying polarization.

**The Dry Cell.** The so-called **dry cell** is really a special type of the early voltaic cell. The electrolyte is nonaqueous and is formed of a paste instead of a liquid, from which characteristic the cell derives its name. This nonspillable feature of the dry cell greatly increases its practicability. Another innovation of the dry cell is the zinc container, which fulfills the double function of acting as the negative electrode and housing the remaining components of the cell.

A dry cell is composed of a positive electrode in the form of a carbon rod immersed in a paste solution contained in a zinc holder which, as mentioned above, forms the negative electrode. The paste is a moist mixture of ammonium chloride, zinc chloride, zinc oxide, plaster of Paris, and sawdust. Usually peroxide of manganese is included in the paste to prevent polarization. The carbon rod is in actual contact with

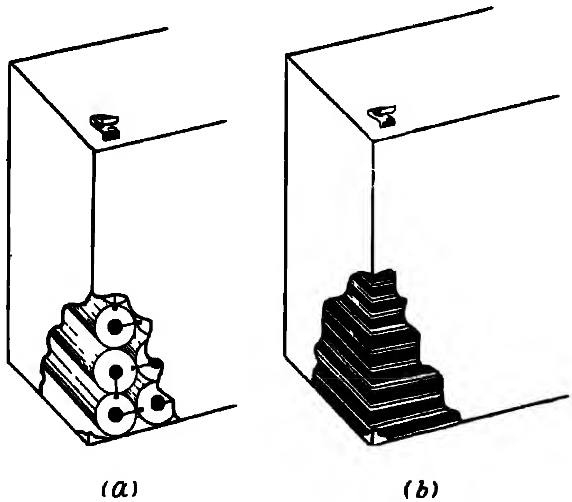


FIG. 18. Construction of dry cell B batteries. (a) Individual-cell construction. (b) Layerbilt construction.

the paste, but the zinc container is separated from it by moistened blotting paper. The entire unit is tightly sealed with sealing wax or pitch to prevent evaporation. Figure 17 shows a cross section of a typical dry cell.

The voltage developed by a dry cell is approximately $1\frac{1}{2}$ v. Dry cells find their greatest application in radio work as B batteries (see Chap. X). They are used wherever it is impractical to obtain d-c power by other means, such as in portable radio equipment, marine radio installations, and emergency equipment. Such batteries consist of a number of dry cells connected in series in order to obtain a high voltage output. A cutaway view of a typical commercial B battery is shown in Fig. 18.

Electric cells are said to be **series-connected** when their terminals are connected in tandem, as shown in Fig. 19(a). In this arrangement, the negative pole of a cell is connected to the positive pole of the adjoining cell. The negative pole of the second cell is connected to the positive pole of the next cell, and so on. The total voltage of a battery of cells series-connected is equal to the sum of the voltages of the individual cells. Thus, if 10 cells of $1\frac{1}{2}$ v each are connected in series, the total voltage of the battery is 15 v.

A number of cells are said to be connected in **parallel, or shunt,** when all the positive terminals of the cells are connected together, and all the negative terminals are connected. This arrangement is shown in Fig. 19(b). With this connection, the total voltage, as measured across the terminals of any of the cells, is the same as the voltage of any of the individual cells, assuming that all the cells are of the same type. Thus, if ten $1\frac{1}{2}$-v dry cells are connected in parallel, the total voltage is $1\frac{1}{2}$ v. The advantage of such an arrangement lies in the fact that the current drawn from each cell in such a battery by a given load is one tenth the current which would be drawn from a single cell connected to the same load. Therefore 10 cells in parallel will last 10 times as long as one cell operating with the same external circuit load.

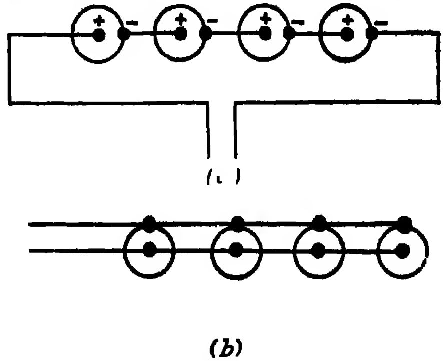



FIG. 19. (a) Dry cells connected in series. (b) Dry cells connected in parallel

### *THE SECONDARY CELL*

There are two principal types of secondary cells in common use today. Both find wide application in radio work and the reader should thoroughly familiarize himself with them. They are the lead acid cell and the Edison cell. Since these cells can be charged by passing an electric current through them, they are often called **storage cells.** Although actually the entire process of charge and discharge is chemical in nature, it is often erroneously thought that electricity is *stored* in secondary cells. This is not true, as will be seen presently. Nevertheless, the action of such cells is sufficiently similar to permit the use of the name "storage cell" or "storage battery" (when referring to a battery of cells) as a matter of convenience.

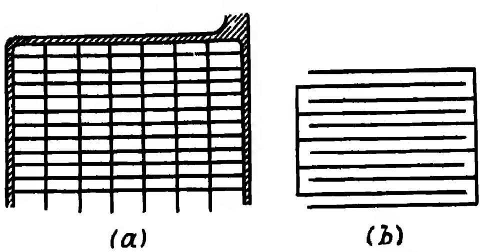



FIG. 20 Construction details of lead acid cell (a) Section of lead-acid plate showing grid construction (b) Method of interposing plates

**The Lead-Acid Cell.** A lead-acid cell, in general consists of lead plates immersed in a dilute solution of sulphuric acid. In a commercial cell, the plates are formed of lead cast into grids, as shown in Fig. 20(a). The

spaces in the grids are then packed with a paste of litharge or oxide of lead. This is the active material of each plate and hardens, or sets, in the grid form like cement. By an electrochemical process, the active material in the positive plate is converted into peroxide of lead when a charging current flows. The active material in the negative plate becomes sponge lead.

A commercial lead-acid cell is made up of a number of positive plates connected by a lead extension, or lug, and a number of negative plates

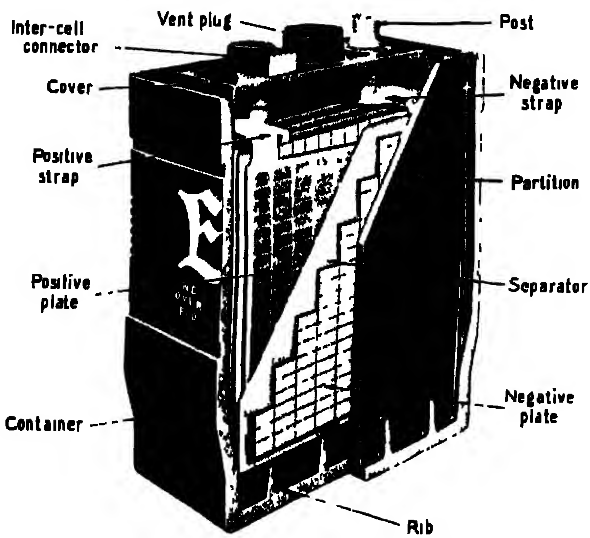



FIG. 21. Cutaway view of lead acid cell. (Courtesy of Electric Storage Battery Co.)

connected in the same manner. The plates are then interposed, as shown in Fig. 20(b), so that the entire unit is composed of alternate positive and negative plates. Commercial cells are customarily constructed with one more negative plate than positive, both outside plates are therefore negative. Although the plates are rigidly mounted porous wood separators are introduced between each pair of plates to prevent contact. To prevent the highly oxidized positive plate from charring the wood, a thin sheet of perforated hard rubber is inserted between the positive plate and the wood separator.

The entire positive and negative plate structure is immersed in a dilute solution of sulphuric acid in a hard-rubber container. Usually, the container is fitted with compartments, permitting the mounting of three cells with separate electrolytes in a common container to form a battery. Such a battery always has its cells series-connected. Figure 21 is a cutaway view of a typical lead-acid cell.

When a lead-acid cell is fully charged, the negative plate is pure

sponge lead (Pb), and the positive plate is lead peroxide ($PbO_2$). If the cell is housed in a glass jar instead of in a rubber container, as many are, the polarity of the plates can be determined by the color of the plates, since lead sponge is light gray in color and lead peroxide is brown. The electrolyte, as mentioned above, is composed of sulphuric acid ($H_2SO_4$) and water ($H_2O$). When the cell is put on discharge, the acid divides into its components $H_2$ and $SO_4$. The $SO_4$ combines with both the positive and negative plates. At the negative plate, it combines with the sponge lead (Pb) to form lead sulphate ($PbSO_4$). At the positive plate, the active material, lead peroxide ($PbO_2$), divides into its component elements lead and oxygen. The $SO_4$ released from the acid combines with the lead released from this plate to again form lead sulphate ($PbSO_4$). The hydrogen released from the acid combines with the oxygen released from the positive plate to form water ($H_2O$) and thus further dilutes the electrolyte. As the discharge progresses, both plates gradually acquire a coating of lead sulphate. The accumulation of lead sulphate eventually fills the pores of the plates and prevents the free circulation of acid about the active material. The action is cumulative, and if the discharge is continued long enough, the cell voltage falls off very rapidly. In practice, discharge is always stopped before the plates become entirely reduced to lead sulphate.

When a cell is charged current is passed through it in a direction opposite to that of discharge. Likewise, the chemical action is the reverse of that in discharge. The lead sulphate ($PbSO_4$) divides into its components Pb and $SO_4$, and electrolysis of the water component of the electrolyte occurs, liberating hydrogen and oxygen. The $SO_4$ released at both plates combines with the freed hydrogen to form sulphuric acid ($H_2SO_4$). The lead (Pb) released at the positive plate combines with the liberated oxygen again to form peroxide of lead ($PbO_2$) at that plate; the lead released at the negative plate is retained at that plate in its original form, that is, sponge lead.

It is apparent that as the cell is charged, the electrolyte becomes increasingly stronger, that is, its acid content becomes greater. Conversely, as the cell is discharged, the electrolyte becomes weaker, that is, its acid content becomes smaller. It follows, therefore, that the relative acid content of the electrolyte is a direct indication of the state of charge of the cell. The acid content of the electrolyte is best ascertained by measuring its specific gravity. The specific gravity of a fully charged lead-acid cell is 1.300. Some types of marine lead acid batteries have lower specific-gravity readings. The specific gravity of a discharged lead-acid cell is 1.100. Since the specific gravity of pure water is 1.000, it is apparent that even when a lead-acid cell is discharged, there is an appreciable acid content in the electrolyte.

The specific gravity of an electrolyte is measured by an instrument

called the **hydrometer** which consists of a glass tube with a small rubber tube at one end and a syringe bulb at the other and inside the tube a small weighted float. Electrolyte is drawn into the device by immersing the rubber tube in the solution, compressing the syringe, and allowing it to expand slowly. The instrument is held in a vertical position, and as the syringe expands, the liquid is drawn into the tube by suction. The weighted float is calibrated so that the depth at which it floats within the glass tube varies with the density (specific gravity) of the liquid in which it is floating. A graduated scale on the float permits the specific gravity to be read directly from the level of the electrolyte about the float.

Great care should be exercised by the operator in handling the electrolyte of lead-acid cells. Severe burns may result if sulphuric acid is allowed to splash upon the operator's person or clothing. The remedy for such acid burns is immediately to apply ammonium hydroxide (common household ammonia). This neutralizes the acid and prevents it from doing further damage. Caution should also be exercised in mixing electrolyte for new installations of lead-acid cells. Sulphuric acid is usually shipped in the concentrated form, and in mixing it with water, it must be remembered that *water should never be added to the acid*. The chemical reaction resulting from this operation is very violent, and there is danger that the resulting steam may splash acid upon the operator. Ordinarily, the manufacturers provide specific instructions for preparing the electrolyte of a lead-acid cell. The standard electrolyte utilizes one part of acid to four parts of water.

**Maintenance of Lead-Acid Cells.** Any impurities in the electrolyte of a lead acid cell may cause it to operate erratically and may permanently damage the cell. Water is continually being lost from the electrolyte of a lead acid cell through evaporation and through excess electrolysis caused by periodic charging. Care should be exercised to inspect lead-acid cells periodically and to replenish the water in the electrolyte. Since the acid neither evaporates nor is consumed, it need not be replenished. The only way acid is lost from a cell is through leakage of a broken container. Should this occur, the plates should immediately be removed from the container and temporarily immersed in water in a nonmetallic container, such as a wooden bucket. Under no circumstances should the plates be allowed to become dry. As soon as possible the plates should be reinstalled in a new container and an entirely new electrolyte added.

Only pure water should be used to replenish the electrolyte of a cell. Distilled water is recommended. Water that has merely been boiled is not satisfactory. Distilled water kept for use in lead-acid cells should never be kept in a metallic container other than lead. Glass, earthenware, hard-rubber, or wooden vessels are commonly used for this purpose.

Containers designed to house the plates of lead-acid cells are customarily constructed with vertical ribs in the bottom of them, as shown in

Fig. 22. The plates of large heavy-duty cells are constructed with feet on the bottom of the plates, which permit the plates to rest on the ribbed bottom of the container. This arrangement allows splinters of wood from the separators, bits of active material washed out of the plates, and other debris that may accumulate in the cell to fall between the ribs to the bottom of the cell. The material is thus prevented from becoming clogged between the plates and causing possible short circuits within the cell. Occasionally in old cells the debris accumulates to the tops of the ribs, causing partial short circuit. When this occurs, the plates should be removed the container thoroughly cleaned, the cell reassembled and entirely new electrolyte added. It is a good plan in such cases also to install new wooden separators, since much of the debris is caused by disintegration of the old separators.

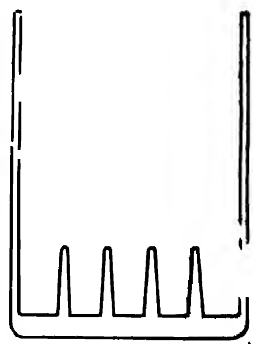


FIG. 22. Cross section of lead acid cell container showing supporting ribs.

As mentioned above, when a lead-acid cell is charged, electrolysis of the water in the electrolyte occurs, decomposing the water into its components hydrogen and oxygen. A considerable portion of these gases unavoidably escapes into the air, and the gases are evidenced by bubbling of the electrolyte Since hydrogen is a highly inflammable gas, caution must be exercised to avoid bringing open flames into the vicinity of charging cells. This liberation of gases from a charging cell is called **gassing.**

A cell can be charged at a high rate until it begins to gas. As the charge progresses, the plates can no longer absorb current at the same rate If the rate of charge is the same, the excess current acts solely to cause electrolysis of the water, that is, to form hydrogen and oxygen gas. This results in excessively high temperature of the cell which may cause the plates to buckle with resultant physical damage to the cell. In addition, such overcharging, if persistent, tends to churn the active material out of the positive plates, thus greatly decreasing the efficiency of the cell.

One of the most harmful actions that occur in a lead acid cell is known as **local action.** Impurities are always unavoidably present in the plates and sometimes in the electrolyte of a cell. These minute impurities act as one electrode of a tiny primary cell, and in conjunction with the plate proper acting as the other electrode small currents are set up within the cell. These local currents reduce the effectiveness of the plates of a cell for their intended purpose, and, in addition, the chemical action produced by them causes the formation of lead sulphate on the plates. The lead sulphate so formed is of a much harder texture than that caused

by normal cell action. Local action is in progress continuously in any lead-acid cell, whether the cell is in use or not; it occurs, regardless of the state of charge of the cell.

If the lead sulphate caused by local action is allowed to accumulate on the plates of a cell, it will increase the internal resistance of the cell to such an extent that the capacity is reduced. Since this sulphate is considerably greater in volume than the active material of the plates, if unchecked, it may cause the plates to buckle and may render the cell completely useless.

Local action is most injurious when the state of charge of a lead-acid cell is low and the temperature is high. Although it cannot be completely eliminated, the effects of local action can be materially decreased by maintaining the cell in a fairly good condition of charge and keeping the temperature as low as possible. Overcharging, with its accompanying high temperature and other detrimental effects, should always be avoided. The lead sulphate produced by local action can sometimes be removed from the plates by subjecting the cell to long low rate chargings.

The plates of a lead acid cell should never be allowed to become dry. Therefore, the electrolyte must always be kept at a height that will completely cover the plates. In most cells, it is customary to keep the level of the electrolyte about ½ in. above the top of the plates. This practice varies with the type and construction of the cell, however, and the operator should be guided by the manufacturer's specifications in each case.

The standard lead-acid cell will provide an emf of 2 v. A cell in good condition will maintain its voltage for a considerable period of time during discharge before beginning to fall off. When the voltage begins to fall off, it is an indication that the lead sulphate is beginning to clog the active material of the plates. If the discharge is continued to an extreme stage, the formation of sulphate may become so great as to injure the cell permanently. To prevent overdischarge, each commercial lead-acid cell has a *discharge voltage limit*, beyond which further discharge may harm the cell. This limit is fixed by the manufacturer and is specified on the name plate of a battery of cells. The discharge-voltage limit varies with different cells, depending upon the type of construction, concentration of the electrolyte, and normal rate of discharge.

The capacity of a secondary cell is usually rated in **ampere-hours,** a unit designated to represent one ampere of current flowing through a given circuit for one hour of time. The ampere-hour capacity of a lead-acid cell depends upon the amount of active surface of the plates exposed to the electrolyte. A cell having a great number of plates will therefore have a correspondingly large ampere-hour capacity. The voltage of such a cell, however, will not exceed the voltage of any other lead-acid cell, regardless of the number of plates.

The manufacturers of lead-acid cells customarily indicate the *normal discharge rate* of a cell. This discharge rate is given in amperes. A given cell will not operate at maximum efficiency if discharged at a rate higher than its normal discharge rate. Thus, a battery of lead-acid cells with a 240 amp-hr capacity and a normal discharge rate of 20 amp will discharge continuously for 12 hr at its normal discharge rate of 20 amp without falling below its discharge voltage limit. Similarly, this battery, if discharged at a 5 amp rate, can be expected to last 48 hr before falling below its discharge-voltage limit. If the battery is discharged at 30 amp, however (*above* its normal discharge rate), it will fall below its discharge voltage limit in considerably less than 8 hr.

Manufacturers also customarily indicate the rate at which a lead acid cell should be charged for maximum efficiency, and care should be taken to abide by the manufacturers' recommendations in all cases. Consistently overcharging a cell will greatly shorten its life.

**The Edison Cell.** The **Edison cell**, also known as the **nickel-iron-alkaline cell**, differs in electrical characteristics, construction, and chemical makeup from any other type of cell and is made only by the Edison Storage Battery Division of Thomas A. Edison, Inc. The Edison cell is an exceedingly long-lived, rugged, and efficient type of secondary cell. It is ideally suited to the type of service required of heavy-duty marine batteries, but because of its greater initial cost, its use is perhaps not so widespread as that of other types.

In general, an Edison cell consists of steel plates packed with active material immersed in a potash electrolyte within a steel container. The positive plate of an Edison cell is composed of a number of perforated steel tubes into which have been packed under heavy pressure alternate layers of nickel hydrate and extremely thin flakes of pure nickel. The negative plate is composed of a steel grid somewhat similar to the grid structure of a lead-acid cell, except that the rectangular grid openings are much larger. Perforated nickel plated steel pockets into which iron oxide has been packed under heavy pressure are supported in the grid openings.

As in the lead-acid cell, a typical Edison cell is composed of a number of positive and negative plates alternately interposed. The plates are prevented from touching each other by rigid, specially treated hard-rubber frames that are mounted at right angles to the plane of the plates. The combined plate structure is mounted in a welded corrugated steel container in an electrolyte consisting of a 21 per cent solution of potassium hydroxide mixed with a small amount of lithium hydroxide in distilled water. All the positive plates and all the negative plates are connected and terminate in steel posts.

When an Edison cell is fully discharged, the active material of the positive plate is nickel oxide (NiO). The active material of the negative

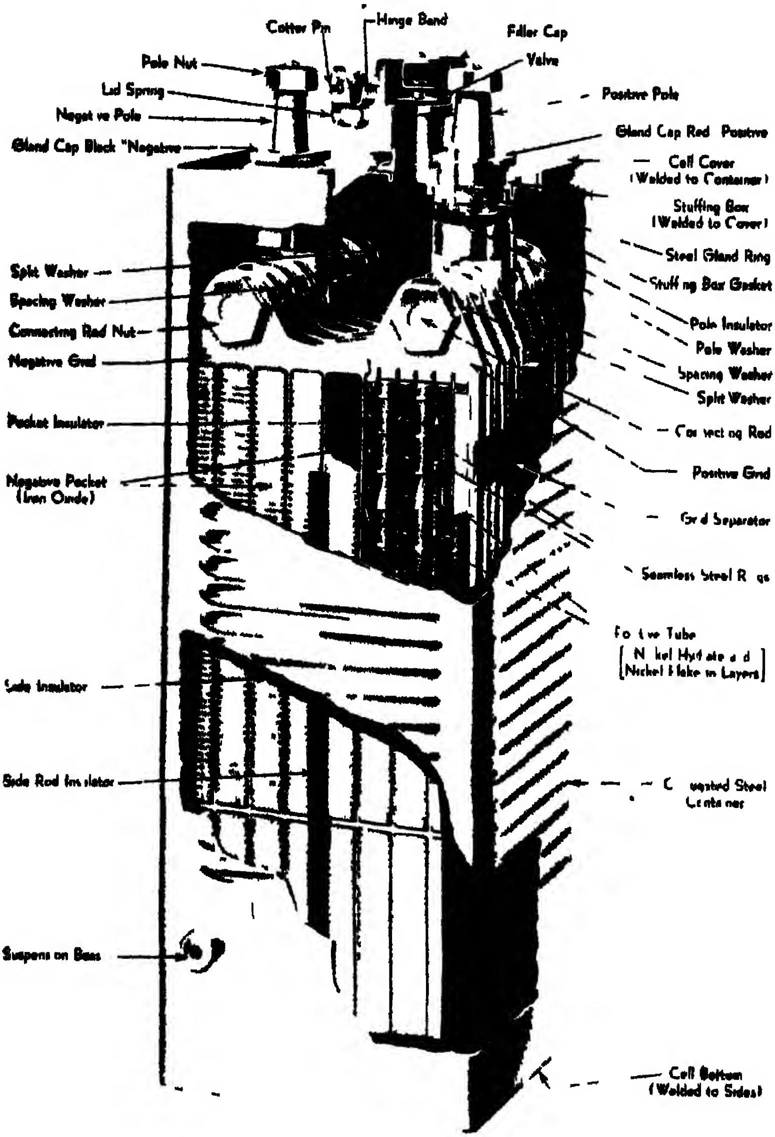



FIG 23 Cutaway view of an Edison cell (*Courtesy of Thomas A Edison, Inc*)

plate is iron oxide ($Fe_3O_4$). The electrolyte, composed of potassium hydroxide (KOH) and water ($H_2O$), appears to act as a catalytic agent during both charge and discharge. When the cell is charged, the iron oxide of the negative plate is reduced to pure iron (Fe). The oxygen liberated from this plate passes to the positive plate where it combines with the nickel oxide (NiO) to form a higher oxide of nickel ($NiO_2$).

During discharge, the action is reversed. The iron of the negative plate is oxidized by the oxygen released from the positive plate as the oxide of nickel in this plate is reduced to a lower oxide. The transfer of oxygen from the positive to the negative plate during this process produces a difference of potential.

The density of the electrolyte during either charge or discharge of an Edison cell remains practically the same. It is apparent, therefore, that the state of charge of an Edison cell cannot be determined by specific-gravity readings of the electrolyte. The manner of checking the state of charge of an Edison cell is discussed later in this chapter.

**Maintenance of Edison Cells.** Since the Edison cell is not subject to such detrimental processes as local action and sulphation, which affect acid cells, its maintenance is comparatively simple. As with the lead-acid cell, it is important that the plates of Edison cells be kept immersed in the electrolyte at all times. This is accomplished by adding distilled water at periodic intervals to replace that lost by electrolysis or evaporation. The cells should be kept clean at all times to prevent the accumulation of potash salts crystals about the vent caps (sometimes called "creeping"). An application of oil or liquid petroleum jelly to the cell tops is efficacious in retarding creeping. The tops of new cells are usually coated with a rosin jelly compounded by the manufacturers for this purpose.

**Secondary-Cell Installation Considerations.** Secondary cells of both the lead-acid and Edison types find wide application as sources of emergency power aboard ships. Ships aboard which a radio station is compulsory under international regulations are required by law to have an emergency power source. The emergency power supply must be capable of delivering sufficient power to operate all emergency radio equipment for a period of not less than 1 hr. On most ships, this emergency source is also utilized for lighting when the regular ship's power has failed. On such ships, emergency power is supplied by a bank of storage batteries, either lead-acid or Edison cells, or by a gasoline- or diesel-driven generator.

The normal voltage in ship electrical distribution systems is 110 v direct current. Aboard large ships, where the emergency power source is designed to carry both the radio station and the ship's emergency-lighting load, a bank of storage batteries is provided which will deliver 120 v. The overvoltage provides for the voltage drop that is to be expected when the battery is connected to a load.

In order that a bank of emergency batteries may be in condition at all times to comply with the minimum power requirements of the law, it is necessary that they be charged often and that periodic maintenance inspections be made. To facilitate the general care of such batteries, most ships provide a separate battery room to house these important units. The battery room should be kept as immaculate by the operator as the operating room itself. The batteries should be greased or oiled regularly (both Edison and lead-acid types) to prevent creeping and the resultant poor electrical connections. Any spilled electrolyte should immediately be neutralized (if a lead acid cell) by application of ammonium hydroxide, as described above. Spilled acid will eat into the wooden trays customarily used to support battery banks. In a remarkably short time, a portion of a wooden battery rack can completely disintegrate because of the action of acid. Care should also be exercised to maintain the battery room watertight, especially in rough weather. If salt water finds its way into the electrolyte of a cell, it may result in permanent injury to the cell. At the same time, it should always be remembered that the battery room must be well ventilated when cells are under charge to permit the exhaust of inflammable hydrogen gas.

Since emergency battery installations are subject to numerous periodic chargings, such installations are usually provided with a so called **charging panel**, which is installed apart from other radio room equipment and provides a flexible means of switching the battery from discharge to charge position and vice versa. Upon the charging panel are also mounted the various safety devices for the protection of the battery and, in some installations, a meter, which indicates the state of charge of the battery.

In order to charge a secondary battery from a given source of direct current, it is necessary that the voltage of the source be higher than the voltage of the battery. Otherwise instead of charging, the battery, having a higher voltage, will discharge into the line. A 120 v bank of storage batteries is therefore divided in two banks of 60 v each. When the batteries are to be charged, the connections are so arranged by switches that the two 60-v banks are connected in parallel across the charging line. When the batteries are to be discharged, the connections are changed by the switches so that the two banks are connected in series, thus providing a discharge voltage of 120 v. The student should study the section on series and parallel circuits in Chap. IV before attempting to analyze charging panel connections.

All charging panels are equipped with fuses for protection against short circuits. Most panels provide fuses in both charge and discharge circuits. Thus, when the battery is on charge, any short circuit in the battery side of the circuit will blow the fuses and prevent overload or other undesirable effects on the ship's generators. Similarly, when the

battery is on discharge, any short circuit in the load circuit will blow the fuses, preventing excessive current from being drawn from the batteries.

A number of charging-panel installations provide circuit breakers in addition to fuses for the protection of the batteries. A circuit breaker is a device which utilizes the principle of electromagnetic induction to open the circuit either on underload or overload. As used in a charging circuit, a circuit breaker acts as an underload device. *The charging current passes through the windings of an iron-core coil (an electromagnet). An iron core, called the* **armature,** *is attracted to the electromagnet by the magnetic field generated as a result of the current flowing through the* windings. When the armature is in this position, a contact on the armature engages a stationary contact mounted on the breaker frame. These contacts are in series with the charging circuit. If, for any reason, the engine room power fails, the electromagnet is de-energized, releasing the armature. When the armature is no longer attracted to the electromagnet, it is repelled by the action of a spring, thus opening the contact. With the circuit thus opened, the batteries are prevented from discharging into the line. When the circuit breaker operates in this manner, it is said to have "tripped." Of course, the breaker must be set by hand originally to allow the holding coil to become energized. Similarly, such a breaker must always be reset by hand after it has tripped. A circuit breaker should never be reset without first ascertaining and correcting the cause of tripping.

Overload circuit breakers are often provided in charging panel installations to prevent the battery from being discharged beyond its normal discharge rate. Such devices ensure that the battery is maintained above its normal discharge-voltage limit for a maximum period of time for a given load. An overload circuit breaker utilizes a contact arm which is closed manually. The contacts are held in this closed position by an iron latch. An electromagnet with its windings in series with the discharge circuit and the contacts is so arranged that its field acts to attract the iron latch. When the current through this coil rises above a certain value, the magnetic field becomes strong enough to attract the latch to the coil. This action releases the contact arm, which falls back, motivated either by gravity or a spring, and breaks the circuit. Provision is usually made to vary the adjustment of the device so that the critical operating current can be varied over a comparatively wide range.

A popular method of maintaining emergency batteries in a charged condition is known as **trickle, or floating, charge.** This method consists of keeping the batteries *continually* on charge at a very low rate and will keep batteries in good condition with a minimum of attention. The operator should see that sufficient ventilation is provided at all times. Periodic inspections must be made to ensure that the level of the electrolyte is maintained at a point that keeps the battery plates immersed at

all times. The recommended rate of trickle charge is usually specified by the manufacturer for a particular battery. In the absence of such a specific recommendation, the proper trickle charge can be determined experimentally. Since this system keeps the battery on charge 24 hr a day, overcharge or undercharge can seriously affect the battery's efficiency. In general, it should be remembered that if the batteries gas continually, they are being overcharged. If the specific gravity (in the case of lead-acid cells) continues to drop, the batteries are being undercharged.

Many charging-panel installations provide trickle-charge circuits. The normal (high charge) circuit includes a number of fixed, and sometimes variable, tapped resistors in series with the line, and the rate of charge may be varied by cutting resistance in or out of the line. Trickle charge is often provided for by a tap on the main charging-resistance network. The necessary additional fixed resistors are then usually permanently connected in the switching circuit. When a battery is kept on trickle charge, it is often said to be "floating across the line."

An instrument often included in standard charging panel installations is the **ampere-hour meter.** This unit is used to indicate indirectly the state of charge of a battery and is of particular advantage in Edison-cell installations. The ampere-hour meter employs a small motor connected in series with the charge and discharge circuits of the battery. The speed at which the motor runs depends upon the amplitude of the current in the circuit. The motor is reversible. When the battery is charging, the motor runs in one direction. When it is discharging, it runs in the opposite direction. The unit is so constructed that rotation of the motor operates a pointer that moves over a dial calibrated in ampere-hours. If the battery is discharged, the pointer moves to the right. If the battery is charged, the pointer moves to the left. The position of the pointer on the dial, therefore, indicates at any instant, the approximate available capacity of the battery. A manually movable red pointer is provided with the unit which is set at the dial setting which coincides with the capacity limit of the battery. The movable motor pointer should never be allowed to go to the right of the red pointer. Whenever the movable pointer reaches the red pointer, it is an indication that the maximum usable capacity of the battery has been utilized and the battery should be recharged. When placed on charge, the charging current causes the pointer to move to the left at a rate determined by the charging current. The ampere-hour meter is provided with a set of contacts designed to prevent overcharge of the battery. As the pointer moves leftward during charge, it encounters a contact just before zero (fully charged position) is reached. The contact so made causes a circuit breaker to trip, disconnecting the battery from the charging source.

On many small ships, a small auxiliary transmitter is utilized for emergency purposes. Such transmitters are customarily powered by a

motor generator operating from a 12-v lead-acid battery. In this arrangement small ships differ from larger ships on which the main transmitter is simply switched over to large battery banks to provide emergency communication. In such small installations, maintenance of emergency batteries is accomplished on a much smaller scale than in the larger installations. The salient features of maintenance are the same, however. A typical 12-v emergency battery with accompanying charging panel is shown in Fig. 24.

Lead-acid batteries are also customarily employed to supply filament power for ship radio receivers, two separate batteries usually being provided for such service. A charging panel is used for receiver batteries which permits one battery to be charged while the other is being used, or discharged, by means of a flexible switching arrangement.

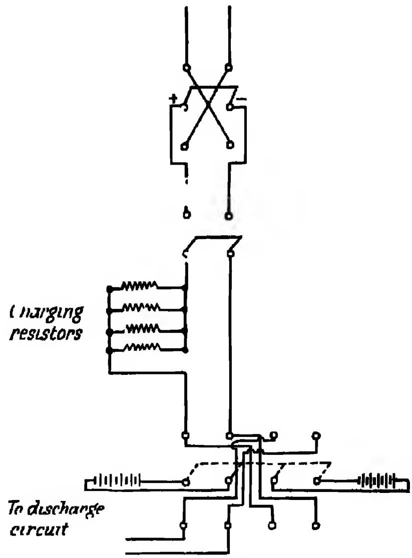



Fig. 24. Typical 12-v charging-panel circuit.

*QUESTIONS AND PROBLEMS**

**1.** What method of connection should be used to obtain the maximum no-load output voltage from a group of similar cells in a storage battery?

**2.** How does a primary cell differ from a secondary cell?

**3.** What is the chemical composition of the active material composing the negative plate of an Edison-type storage cell?

**4.** What is polarization as applied to a primary cell, and how may its effect be counteracted?

**5.** How should sulphuric acid and water be mixed if it becomes necessary to do so in order to replace lost electrolyte?

**6.** What is meant by the term "sulphation" as applied to a lead-acid storage cell?

**7.** What is the chemical composition of the active material composing the positive plate of a lead-acid type storage cell?

**8.** What is the chemical composition of the electrolyte used in an Edison-type storage cell?

**9.** What is the chemical composition of the electrolyte in a lead-acid storage cell?

**10.** What method of connection should be used to obtain the maximum short-circuit current from a group of similar cells in a storage battery?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

# Chapter IV

# DIRECT-CURRENT THEORY

In d-c work there are two basic types of networks to be considered—the series circuit and the parallel circuit. The study of d-c theory is primarily the study of current and voltage relations in series and parallel circuits and the calculation of the resistance of such circuits. All d-c networks, no matter how complex, can be reduced to combinations of series or parallel circuits or both. Once the principle of series and parallel circuits is understood, the solution of any d-c network is simply a matter of breaking down the network into the component fundamental circuits.

## *OHM'S LAW*

In 1827, when Georg Simon Ohm formulated the famous law named after him, one of the most important forward steps in the history of electrical science was made. His law laid the groundwork for practically all our electrical theory as we know it today. The student should not underestimate the importance of Ohm's law, despite its apparent simplicity at first glance. Thorough understanding and intelligent application of this law are absolutely essential for the proper solution of *every* major circuit used in radio work. Ohm's law states:

*The current in amperes in any given circuit is directly proportional to the electromotive force in volts and inversely proportional to the resistance in ohms.*

Notice that all values in Ohm's law are given in units, that is, amperes, volts, and ohms. Care must be taken to reduce all values to units before applying Ohm's law to a given problem.

**Mathematical Forms of Ohm's Law.** Ohm's law, as it is given above, is a *statement* of the relation that exists between current, voltage, and resistance in a circuit. It is more convenient in applying this law to represent this relationship mathematically as a formula. Presented mathematically, Ohm's law states

$$I = \frac{E}{R}, \tag{1}$$

where $I$ = standard symbol for current;
$E$ = standard symbol for voltage;
$R$ = standard symbol for resistance.

In Chap. II we learned that an equation in the form of Eq. (1) is a simple linear equation. We learned, too, that we could solve such an equation for any unknown in terms of the remaining values by simply clearing of fractions and making the proper transposition. In Eq. (1) we have $I$ in terms of $E$ and $R$. In order to find $E$ in terms of $I$ and $R$, we simply clear Eq. (1) of fractions by multiplying both sides by $R$; thus,

$$IR = E, \tag{2}$$

or, by rearranging as it is more generally used,

$$E = IR. \tag{3}$$

To find $R$ in terms of $E$ and $I$, we divide both sides of Eq. (2) by $I$, giving us

$$R = \frac{E}{I} \tag{4}$$

We find therefore, that the statement of Ohm's law can be arranged mathematically in three ways, that is, Eqs. (1), (3), and (4). The student, of course, need remember only one of these forms. By applying algebra, it is a simple matter to derive each of the remaining forms.

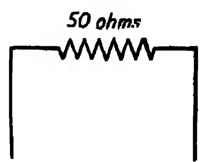



FIG. 25.

**Problem.** If an emf of 100 v is applied across the resistor shown in Fig. 25, what current will flow through the resistor?

**Solution.** Since the unknown quantity in this case is the current, or $I$, it is convenient to apply Ohm's law in the form of Eq. (1). Thus,

$$I = \frac{E}{R} \tag{1}$$

Substituting the known values of $E$ and $R$,

$$I = \frac{100}{50}, \tag{5}$$

or

$$I = 2. \tag{6}$$

Since the known values of $E$ and $R$ were given in units, the answer according to Ohm's law will also be in units, or

$$I = 2 \text{ amp.} \tag{7}$$

**Problem.** If the value of the resistance shown in Fig. 25 is unknown, how can we ascertain this value if it is known that an emf of 100 v applied across the resistor will produce a current flow of 25 ma?

**Solution.** In this problem, we are given the values of $E$ and $I$, but $R$ is unknown. It is apparent, therefore, that Eq. (4) is the proper form of Ohm's law to use and since Ohm's law is stated in units, we must first reduce our

given values to units before applying Eq. (4). The voltage is already given in units and may be applied directly. The current, however, must first be reduced to amperes.

According to the conversion table given in Chap. 1, the prefix *milli* indicates *one thousandth*. Making use of exponents, we find,

$$25 \text{ ma} = 25 \cdot 10^{-3} \text{ amp}, \tag{8}$$

or

$$25 \text{ ma} = 0.025 \text{ amp}. \tag{9}$$

Since both given values have now been reduced to units, we may proceed in applying Ohm's law.

$$R = \frac{E}{I}. \tag{4}$$

Substituting,

$$R = \frac{100}{0.025}. \tag{10}$$

Solving,

$$R = 4{,}000 \text{ ohms} \tag{11}$$

**Power in D-C Circuits.** The unit of electric power, as we know, is the **watt.** According to the basic definition of the watt, as given in Chap. I, one watt represents the rate at which work is done by a current of one ampere at a pressure of one volt; or, expressed mathematically,

$$1 \text{ w} = 1 \text{ v} \cdot 1 \text{ amp}. \tag{12}$$

Stating Eq. (12) as a general formula, we may say

$$P = EI, \tag{13}$$

where $P$ = power in watts;
$E$ = electromotive force in volts;
$I$ = current in amperes.

The power in any d-c circuit can thus be obtained by substituting the values of voltage and current of the circuit in Eq. (13) and solving for $P$. It will be noted that here again, all given values must be reduced to units before substitution.

Often it is desirable to ascertain the power expended in a circuit when the voltage and resistance or the current and resistance are the only known values. Power equations have been derived to meet this need by the application of Ohm's law to Eq. (13).

$$P = EI. \tag{13}$$

According to Ohm's law,

$$E = IR. \tag{3}$$

Substituting in Eq. (13) the value of $E$ given in Eq. (3), we get

$$P = IRI, \tag{14}$$

or

$$P = I^2R. \tag{15}$$

Equation (15), therefore, gives us the power in watts when the current in amperes and the resistance in ohms are known.

In like manner, we may substitute a value for $I$ in Eq. (13). Since, according to Ohm's law,

$$I = \frac{E}{R}, \tag{1}$$

we may take this value of $I$ given in Eq. (1) and substitute it in Eq. (13), giving us

$$P = E \cdot \frac{E}{R} \tag{16}$$

or

$$P = \frac{E^2}{R}. \tag{17}$$

We now have three forms of the power equations—Eqs. (13), (15), and (17). All three forms should be memorized, since frequent use of them will be made throughout the text.

In d-c work, we deal constantly with three circuit values, namely, voltage, current, and resistance. When any *two* of these values are known, we can find the third by the application of the proper Ohm's law equation, and we can find the power expended in the circuit by the application of the corresponding power equation. Conversely, if the wattage and any *one* of the above values are known, we can ascertain any or all of the remaining circuit values.

**Problem.** In the 50-ohm resistor illustrated in Fig. 25, 200 w of power is being expended. What is the voltage across this resistor? What current flows through the resistor?

**Solution.**

$$P = I^2R. \tag{15}$$

Dividing both sides by $R$,

$$\frac{P}{R} = I^2. \tag{18}$$

Extracting the square root of both sides,

$$\sqrt{\frac{P}{R}} = I. \tag{19}$$

Rearranging,

$$I = \sqrt{\frac{P}{R}}. \tag{20}$$

Substituting,

$$I = \sqrt{\frac{200}{50}}, \tag{21}$$

or

$$I = \sqrt{4} = 2 \text{ amp}. \tag{22}$$

Then,

$$E = IR. \tag{3}$$

Substituting,

$$E = 2 \cdot 50, \tag{23}$$

$$E = 100 \text{ v.} \tag{24}$$

It should be remembered that Ohm's law and the power equations can be applied not only to an *entire* circuit but also to any *portion* of a circuit. In the latter case it is apparent, of course, that the values substituted in the formulas are true only for the particular portion of the circuit to which the formula is being applied. Applications of these formulas to portions of complex circuits will be illustrated in a later part of this chapter.

FIG. 26. A much enlarged cutaway view of an insulated metallized resistor. Glass tube is coated with a metallized resistance material and connected to external leads. The entire unit is sealed by a molded insulating phenolic. (*Courtesy of International Resistance Co.*)

**A resistor,** as the term is used in electric circuits, is a conductor in which is concentrated, or lumped, a comparatively large amount of actual ohmic resistance to the flow of an electric current. Another way of describing it would be to say that a resistor is a poor conductor. The higher the resistance value of a resistor, the poorer the conductor it is. Commercial resistors are usually made of carbon or other high-resistance material or are composed of many turns of high resistance wire wound about a cylindrical form of insulating material.

Resistors are used in radio circuits wherever it is desirable to *decrease* the amount of electric energy being applied to an electric circuit. For example, when two portions of a circuit are designed to operate at different voltages, it is usually not convenient to utilize two separate sources of voltage. *One* power supply is used, which supplies the higher-voltage portion of the circuit. A resistor of the proper value is then inserted in the circuit between the voltage source and the lower-voltage portion of the circuit, assuring that the proper voltage differential is maintained.

According to the theory of the conservation of energy, energy cannot be destroyed. When energy is utilized to do work of any kind, we do not

*lose* that energy; it is merely being transformed into energy in another form. Thus, when electric energy is expended in a resistor, as in the problems above, it does not disappear but assumes another form. In this instance, the electrical energy is dissipated in the form of *heat*. Heat, although a very useful form of energy for some applications, can also be a destructive factor. Therefore, resistors must be safeguarded against the heat which is dissipated in them when used in electric circuits. All

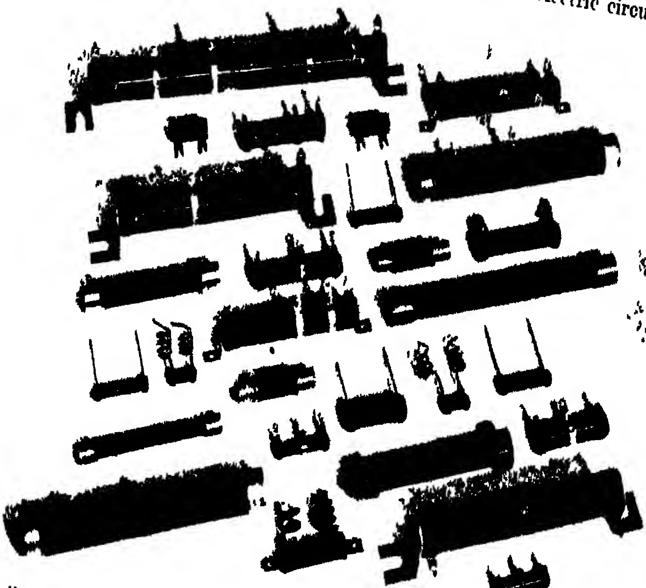


FIG. 27. A number of typical wire wound resistors. (*Courtesy of International Resistance Co.*)

resistors have a certain limit beyond which any additional heat generated by transformation of energy within them will cause them to oxidize, or burn up.

Commercial resistors are rated with this *heat dissipation factor* in mind. Thus, it is necessary not only to stipulate that a resistor used in a given circuit must be of a given ohmic resistance, but also to specify the wattage the resistor will have to dissipate in the form of heat. The resistor, then, will assuredly be able to withstand this heat without ill effects.

Commercial resistors are usually conservatively rated. Nevertheless, it is standard practice to add a *safety factor* of approximately 50 to 100 per cent. This added assurance is necessary since resistors are often so

located in chassis or cabinets that they do not obtain the amount of air circulation essential to cooling. Thus, for the resistor used in the last problem, the actual wattage of the circuit is 200 w. In purchasing a resistor for use in that circuit, standard practice would dictate the purchase of a 50-ohm 300-w resistor. Generally speaking, the smaller the resistance value, the higher the safety factor that should be allowed.

## *SERIES CIRCUITS*

**Characteristics.** There are only two basic ways in which electrical components of a circuit may be connected, namely, in series or in parallel.

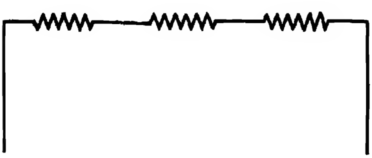


FIG. 28.

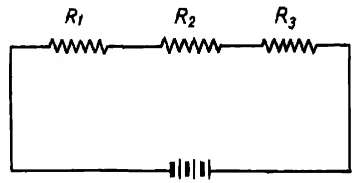



FIG. 29.

When two or more components are connected in tandem, as in Fig. 28, they are said to be **series-connected.**

If the three resistors shown in the series circuit of Fig. 28 are connected to a battery, as shown in Fig. 29, we know that current will flow from the positive pole of the battery through each of the resistors in turn before completing the circuit to the negative pole. If any one of these resistors should become open-circuited, we know that no current would flow through *any* of the resistors, since the only path by which the current can possibly complete the circuit is through each of the resistors in turn. It is evident that the same current flows through all the components in such a circuit. We can therefore formulate a fundamental law for such circuits: *The current is the same in all parts of a series circuit.*

The total resistance of all the resistors in a series circuit is obviously the sum of the individual resistances. Expressed mathematically,

$$R_t = R_1 + R_2 + R_3 + \cdot \cdot \cdot, \tag{25}$$

where $R_t$ = total resistance.

**Ohm's Law in Series Circuits.** If, in Fig. 29, we assume a value of 10 ohms for $R_1$, 20 ohms for $R_2$, and 30 ohms for $R_3$, then, according to Eq. (25), the total resistance, or $R_t$, will equal 60 ohms. If the battery shown in Fig. 29 delivers a voltage of 6 v, then, according to Ohm's law,

$$I = \frac{E}{R}, \tag{1}$$

or,

$$I = \frac{6}{60}, \tag{26}$$

and

$$I = 0.1 \text{ amp.} \tag{27}$$

According to our first rule for series circuits, therefore, we know that 0.1 amp flows through each of the resistors in this circuit. Since we know the resistance in each case we can then apply Ohm's law to each portion of the circuit in turn and find the voltage drop across each resistor

$$E_1 = IR_1. \tag{28}$$

Substituting,

$$E_1 = 0.1 \cdot 10, \tag{29}$$

and

$$E_1 = 1 \text{ v.} \tag{30}$$

Similarly,

$$E_2 = IR_2. \tag{31}$$

Substituting,

$$E_2 = 0.1 \cdot 20, \tag{32}$$

and

$$E_2 = 2 \text{ v.} \tag{33}$$

Also,

$$E_3 = IR_3. \tag{34}$$

Substituting,

$$E_3 = 0.1 \cdot 30,$$

and

$$E_3 = 3 \text{ v} \tag{35}$$

where $E_1$ = voltage drop across $R_1$,
$E_2$ = voltage drop across $R_2$,
$E_3$ = voltage drop across $R_3$

It is at once apparent that there is a certain similarity between the voltages across the components and the resistances of the components. Left to right in Fig. 29, the resistances are 10, 20, and 30 ohms; the voltages 1, 2, and 3 v. The voltages, we note, are proportional to the resistances across which they appear. This is not just a coincidence for this particular case. Had we chosen odd values of resistance, such as 31, 47, and 83 ohms, we would have found that the voltages were exactly proportional to the resistances across which they appear. That this is true for all cases can be realized by an inspection of the Ohm's law equation

$$E = IR. \tag{3}$$

If $I$ remains the same (as it does in all parts of a series circuit), $E$ will vary directly with $R$. In other words, if $R$ is increased, $E$ will increase proportionally. If $R$ is decreased, $E$ will decrease proportionally. As we have seen with the series circuit of Fig. 29, $R_2$ was twice the resistance of $R_1$, consequently, $E_2$ was twice the voltage of $E_1$. If $R$ is tripled, $E$ is tripled, and so on. We can therefore formulate our second fundamental law for series circuits

*The voltage across any part of a series circuit is directly proportional to the resistance of that part of the circuit.*

## PARALLEL CIRCUITS

**Characteristics.** When two or more components of a circuit are connected across each other or across a common voltage source, they are said to be **parallel-,** or **shunt-,** connected. If a battery is connected to the circuit, as shown in Fig. 30, it is apparent that the battery voltage is applied equally across all three resistors. This is obvious since one terminal of *each* resistor is directly connected to a positive pole of the battery and the other terminal of each resistor is directly connected to a negative pole of the battery. It follows, therefore that the first fundamental law for parallel circuits is.

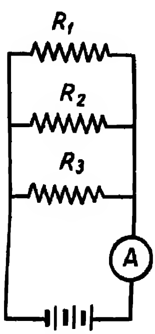


FIG. 30.

*The voltage is the same across all branches of a parallel circuit.*

Computing the total resistance of a parallel circuit is a slightly more complex procedure than for a series circuit. The total resistance of a parallel circuit is equal to the reciprocal of the sum of the reciprocals of the individual resistances. This can more clearly be expressed mathematically as

$$R_t = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots} \tag{36}$$

or, expressed differently,

$$\frac{1}{R_t} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \tag{37}$$

Formulas (36) and (37) apply to any parallel circuit, regardless of the number of branches. There are certain special cases of resistors in parallel, however, for which short cuts are available. Two such cases are: when all the resistors in a parallel network are of equal value and when there are only two branches to a parallel network.

When all the resistors in a parallel network are of equal value, the total resistance is found by simply dividing the value of *one* resistor by the *number* of resistors in the network. Thus, if two 10-ohm resistors are in parallel, the total resistance is 10 divided by 2, or 5 ohms. If six 10-ohm resistors are in parallel, the total resistance is 10 divided by 6, or 1.6 ohms. Expressed as a formula for any number of equal resistors,

$$R_t = \frac{R_1}{N}, \tag{38}$$

where $R_t$ total resistance;
$R_1$ resistance of any one resistor,
$N$ number of resistors.

When there are only two resistors to a parallel network, the total resistance may be found by dividing their product by their sum. This holds true whether or not the resistors are of equal value. Expressed mathematically,

$$R_t = \frac{R_1 \cdot R_2}{R_1 + R_2}. \tag{39}$$

One fact should be remembered in computing parallel-circuit resistance —the total resistance is *always less* than the resistance of the *smallest* component resistor.

**Ohm's Law in Parallel Circuits.** If the resistors in Fig. 30 are assigned values of 10, 20, and 30 ohms, respectively, for $R_1$, $R_2$, and $R_3$, substitution in Eq. (36) will give a total resistance of 5.45 ohms for the circuit. If the battery shown in the diagram delivers 6 v, then, according to Ohm's law,

$$I = \frac{E}{R}. \tag{1}$$

Substituting,

$$I = \frac{6}{5.45}, \tag{40}$$

and

$$I = 1.1 \text{ amp} \tag{41}$$

This current of 1.1 amp is the *total* circuit current, since we have applied Ohm's law to the *entire* circuit, and is the current that would be indicated by the ammeter in Fig. 30. According to our first law for parallel circuits, the voltage is the same across all branches of a parallel circuit. Therefore, we have two known values for each branch of the parallel circuit. We know the voltage and the resistance of each leg. We can then readily apply Ohm's law to each resistor in turn to find the remaining unknown value, the current in each leg. The particular Ohm's law equation which applies in this problem is again

$$I = \frac{E}{R} \tag{1}$$

Substituting,

$$I_1 = \frac{6}{10} = 0.6 \text{ amp}, \tag{42}$$

$$I_2 = \frac{6}{20} = 0.3 \text{ amp}. \tag{43}$$

$$I_3 = \frac{6}{30} = 0.2 \text{ amp}. \tag{44}$$

Here again certain similarities stand out. It is at once apparent that the maximum portion of the total current flows through the smallest resistor. As the resistors become larger, the current flowing through them becomes smaller. When $R$ is increased from 10 to 20 ohms, as illustrated by the resistors $R_1$ and $R_2$ in the example, $I$ is decreased from 0.6 to 0.3 amp ($I_1$ and $I_2$). In other words, when the resistance is doubled, the current is halved. When the resistance is tripled, as illustrated by $R_3$, the current is reduced to one third. That this relation is true for all cases can be seen by an inspection of the Ohm's law equation

$$I = \frac{E}{R}. \tag{1}$$

It is obvious that any increase in $R$ will cause a proportional decrease in the total value of the fraction $\frac{E}{R}$ or $I$; conversely, any decrease in $R$ will cause a proportional increase in $I$.

Another fact disclosed by solution of the above problem is that the sum of the branch currents equals the total current. This statement must obviously be true for all cases, since there is no other way for the current to complete the circuit to the battery other than through the various parallel legs of the circuit.

The second fundamental law for parallel circuits, therefore, is:

*The current through each branch of a parallel circuit is inversely proportional to the resistance through which it flows, and the algebraic sum of the branch currents is equal to the total current.*

## KIRCHHOFF'S LAWS

**Kirchhoff's First Law.** Ohm's law, as we have said before, is all-important. It is a direct application of the law of the conservation of energy to the phenomena of electric currents. Useful as Ohm's law is, however, there are certain complex circuit combinations that cannot be solved by its use alone. First to realize the need for supplementary rules was the German physicist Gustaf Robert Kirchhoff. In the early part of the nineteenth century, Kirchhoff developed a theorem that further

clarified the distribution of currents in a network. He extended Ohm's law for a linear conductor to the case of conductors in three dimensions, in this manner generalizing the equations dealing with the flow of electricity in conductors.

Kirchhoff's theorem has been handed down to us in the form of two basic laws that can be applied directly to specific problems. Kirchhoff's first law states:

*The algebraic sum of the voltage drops around every closed circuit, including the source, is always equal to zero*

This law simply states in effect that no electricity accumulates at any point in a circuit. Another way of stating Kirchhoff's first law is: *The sum of the IR drops in any circuit is equal to the algebraic sum of the various emfs acting in the circuit.*

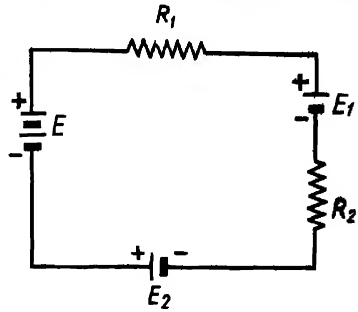


FIG. 31

Proper application of this apparently obvious fact considerably simplifies the solution of many networks. In the circuit in Fig. 31, there are three separate voltage sources, two voltages supplementing each other and the third opposing the two. The circuit is easily solved for any unknown value by application of Kirchhoff's first law. Thus,

$$E + E_2 - E_1 = IR_1 + IR_2. \tag{45}$$

This equation is simply an expression of the known facts about this circuit according to Kirchhoff's law. Nevertheless, Eq. (45) may be used to find any unknown in the circuit by merely solving for the unknown in terms of the remaining known values. If, for example, it is desired to ascertain the potential existing across the resistor $R_1$, we simply make the necessary transpositions in Eq. (45) and obtain

$$IR_1 = E + E_2 - E_1 - IR_2. \tag{46}$$

**Kirchhoff's Second Law.** Kirchhoff's second law states:

*At any point in a circuit there is as much current flowing to the point as there is flowing away from it.*

Another way of expressing this rule is: *The algebraic sum of the currents at any point in a circuit is always zero.*

As an example of the use of Kirchhoff's laws, consider the circuit of Fig. 32, in which the current $I$ and the resistors $R_1$, $R_2$, and so on, are

known, but the five currents $I_1$, $I_2$, and so on, and the five voltages $E_1$, $E_2$, and so on, are unknown. As will be shown, Ohm's law alone is not sufficient for solving a circuit of this type.

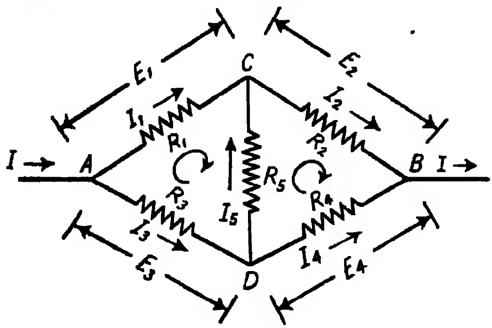


FIG. 32

Ohm's law allows us to write the following five equations for this circuit:

$$E_1 = I_1R_1, \tag{47}$$
$$E_2 = I_2R_2, \tag{48}$$
$$E_3 = I_3R_3, \tag{49}$$
$$E_4 = I_4R_4, \tag{50}$$
$$E_5 = I_5R_5. \tag{51}$$

However, 10 equations are needed to solve the problem as there are 10 unknown quantities. Two more equations are obtained from Kirchhoff's first law. Equating the voltage drop in the loop $DACD$ to zero,

$$-I_3R_3 + I_1R_1 - I_5R_5 = 0. \tag{52}$$

The two minus signs are due to the fact that the direction taken around the loop is opposite to the assumed direction of the currents $I_3$ and $I_5$. (If this assumed direction is incorrect in the case of $I_5$, for example, the value of $I_5$ will turn out to be negative in the final solution.)


FIG. 33. A variable resistor or potentiometer. (*Courtesy of Clarostat Manufacturing Co., Inc.*)

Similarly, for the loop $DCBD$ we obtain

$$I_5R_5 + I_2R_2 - I_4R_4 = 0 \tag{53}$$

The three remaining equations necessary for a solution are obtained by means of the application of Kirchhoff's second law. Equating the current entering to the current leaving point $A$, we have,

$$I = I_1 + I_3. \tag{54}$$

For junctions $B$ and $C$,

$$I_2 + I_4 = I, \tag{55}$$

and

$$I_1 + I_5 = I_2. \tag{56}$$

It is not necessary to write the equation for point $D$, because all the currents entering or leaving this point are included in the equations for

points $A$, $B$, and $C$. We now have 10 simultaneous equations in 10 unknowns, which may be solved by using the methods of Chap. II.

## *D-C METERS*

Electricity can be observed only indirectly by detecting its effects through the use of specially designed instruments. Such instruments enable us to measure accurately the voltage, current, and other electric-circuit values.

**The Ammeter.** Perhaps the simplest instrument used in radio work is the ammeter. It depends for its operation upon the principle of electromagnetic induction discussed in Chap. I. As shown in Fig. 34(a), a coil

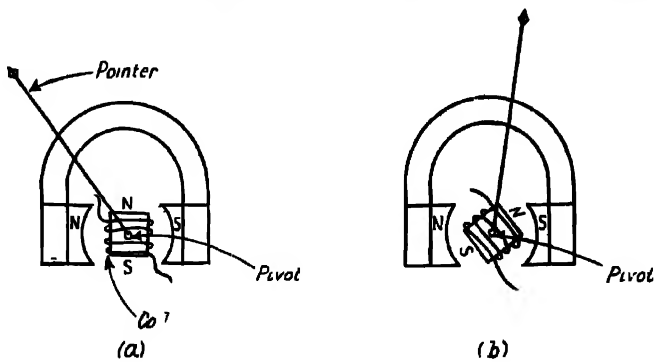



FIG. 34

is mounted upon pivots between the poles of a permanent magnet. Current flowing through the coil will create a magnetic field that is the same as that formed by the north and south poles of a short magnet, giving rise to repulsion and attraction forces between the coil and the permanent magnet. The north pole of the coil will be attracted by the south pole and repelled by the north pole of the permanent magnet. Similarly, the south pole of the coil will be attracted by the north pole and repelled by the south pole of the permanent magnet. These forces will cause the coil to rotate until the electromagnetic forces are balanced by the force of a spiral spring that is mounted so that it tends to return the coil to its neutral position. A pointer is mounted upon the coil so that it moves over a graduated scale as the coil rotates.

Since the magnetic field created by the permanent magnet is a *fixed* value, the amount that the coil rotates about its axis will depend directly upon the strength of the electromagnetic field about the coil. As we know, the strength of the electromagnetic field is directly proportional to the current flowing through the coil. The pointer movement across the graduated scale therefore is a direct indication of the amount of current

flowing through the coil. All that remains is to calibrate the graduated scale in units of amperes. This principle is the basis of the Weston movement, which was developed in the late nineteenth century by Dr. Edward Weston and is used in most d-c instruments today.

**Sensitivity.** The sensitivity of a meter refers to the amount of current necessary to move the pointer across the full scale of the meter dial. Thus, a meter that requires a current of 1 ma for full-scale deflection is said to have a **sensitivity** of 1 ma. If a current meter is calibrated in amperes, it is called an **ammeter**; if calibrated in milliamperes, a **milliammeter**; if in microamperes, a **microammeter**; and so on. Current meters having a sensitivity of 50 $\mu$a are quite common.

Instead of constructing a variety of meters with different sized coils to

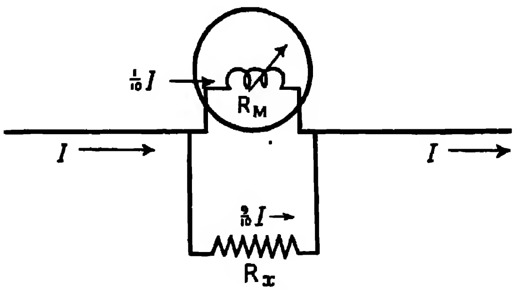



FIG. 35 A typical ammeter shunt circuit

meet various needs, a common practice is to utilize a meter movement of fairly high sensitivity and equip the instrument with one or more *shunts* to extend its range. The utility of a given instrument is thus greatly increased at comparatively small additional cost. A shunt, in this sense, is simply a resistor that is *shunted*, or *paralleled*, across the meter coil (see Fig. 35). If the meter in the diagram has a sensitivity of 1 amp and an internal resistance of 1 ohm, the necessary value of shunt resistance to extend the range of the instrument to any given value can be easily computed Since the circuit is a simple parallel circuit of two branches, the current rule for parallel circuits applies to it

**Problem.** If the ammeter in Fig. 35 has a sensitivity of 1 amp and an internal resistance of 1 ohm, what value of resistance must be shunted across it to extend the range to 10 amp?

**Solution.** According to the current rule for parallel circuits, the current in each branch is inversely proportional to the resistance through which it flows. Therefore,

$$\frac{I_m}{I_x} = \frac{R_x}{R_m}, \tag{57}$$

where $I_m = 1$ amp = current flowing through meter;
$I_x = 9$ amp = current flowing through resistor;
$R_m = 1$ ohm = internal resistance of meter;
$R_x$ = resistance of shunt resistor.

Since the meter has a sensitivity of 1 amp, we cannot allow more than 1 amp to flow through it. The remaining 9 amp of the total current of 10 amp must flow through the shunt. Substituting,

$$\frac{1}{9} = \frac{R_x}{1}, \tag{58}$$

$$9R_x = 1,$$

and

$$R_x = \frac{1}{9} \tag{59}$$

The scale reading in the above illustration must, of course, be multiplied by 10 to obtain the true total-current reading.

It is apparent that a given current meter may be equipped with a number of shunts, thereby greatly increasing the range of the instrument. This practice is common in radio test equipment, the various shunts being connected by means of a multitapped switch.

Any type of *current* meter should always be connected in *series* with the circuit whose current is being measured.

**The Voltmeter.** A d-c voltmeter is really a special application of a current meter. According to the Ohm's law equation

$$I = \frac{E}{R}, \tag{1}$$

if the resistance of a circuit is *constant*, the current flowing through it is directly proportional to the voltage across it. Any given d-c ammeter has a constant value of resistance which is unaffected by the amount of current flowing through it. If, therefore, such an ammeter is connected *across* a circuit instead of in series with it, the current flowing through the meter as indicated on the meter scale is directly proportional to the voltage of the circuit. Consequently, we can calibrate the scale of the meter in *volts* instead of *amperes* and use the meter to indicate voltage.

If an ammeter, such as the one used in the problem above, were connected across a 100-v circuit, the current through the meter, according to Ohm's law, would be 100 amp. Obviously the meter, since it is designed to carry only 1 amp, would burn out. Clearly some resistance must be inserted in the meter circuit to limit the current to the meter capacity. The value of this resistance is easily determined by Ohm's law

$$R = \frac{E}{I} \tag{4}$$

Substituting,

$$R = \frac{100}{1} = 100 \text{ ohms.} \tag{60}$$

In this case 100 ohms is the total resistance of the meter circuit, including the meter internal resistance of 1 ohm. The resistance that must be added, therefore, is 99 ohms. A current of 1 amp flowing through this meter circuit (full-scale deflection) would indicate 100 v across the circuit. Such a voltmeter would be said to have a *sensitivity* of *1 ohm per volt*. Actually, a 1-ohm-per-volt meter would be highly impractical for radio work, since the excessive current (1 amp) drawn by the meter would seriously disturb normal circuit conditions. A *practical* voltmeter has a very high resistance, thereby adding very little load to the circuit under measurement. Most commercial voltmeters have a sensitivity of 1,000 ohms per volt; and meters with sensitivities of 10,000 and 20,000 ohms per volt are quite common.

## RESISTANCE MEASUREMENTS

Circuit values of *voltage* and *current* can be obtained directly by means of the instruments described above. The *resistance* of a circuit, however, cannot be so readily ascertained. There is no instrument that can be connected directly into a live circuit to indicate instantaneous values of circuit resistance; it is therefore necessary to resort to indirect methods to obtain this value. There are numerous ways of measuring resistance, but we shall discuss four of the more commonly used methods.

**Voltmeter-Ammeter Method.** The voltmeter-ammeter method is probably the simplest, although not always the most convenient, to use. The method consists of taking voltage and current readings with the unknown resistance in the circuit and then applying Ohm's law to find $R$.

$$R = \frac{E}{I}. \tag{4}$$

The disadvantages of this system are obvious. It necessitates the use of *two* instruments to obtain *one* circuit value; a separate power supply is required; and the range of resistances that can be measured is dependent upon the types of voltmeter and ammeter that are available.

**Ohmmeter Method.** It has been shown above how a current meter can be adapted to reading voltage through application of the Ohm's law equation

$$I = \frac{E}{R}. \tag{1}$$

*In the voltmeter circuit, it is apparent from Eq. (1) that with $R$ constant any variation in voltage $E$ produces a directly proportional*

variation in current $I$ with a resultant proportional movement of the meter pointer over the scale.

In like manner, if $E$ is kept constant and $R$ is varied, it is evident from an inspection of Eq. (1) that there will be a resultant variation of $I$. In this case, however, it is apparent that *increasing* $R$ will *decrease* $I$, and, conversely, *decreasing* $R$ will *increase* $I$. In other words, $I$ is *inversely proportional* to $R$.

This principle is utilized in adapting current meters to measure values of unknown resistance. A sensitive current meter, usually a 0–1 milliammeter or a microammeter, is calibrated *backward* in ohms. That is, maximum ohms are registered at the zero current position and minimum or zero ohms at the maximum current position. Constant $E$ is maintained by including a small low voltage dry battery in the circuit. The battery is customarily mounted directly in the meter housing. Such a meter is called an **ohmmeter**, and the type just described is a **series-type** ohmmeter.

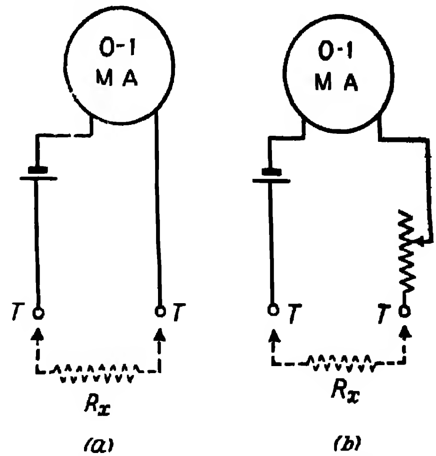


FIG. 36. Fundamental circuits for series-type ohmmeters.

Since $E$ is kept constant in an ohmmeter circuit, some provision must be made to establish a *minimum* value of $R$ in order to safeguard the meter against excessive current. Figure 36(a) is a diagram of a series-type ohmmeter utilizing a 0–1 milliammeter and a 1.5 v dry cell. If a 3-ohm resistor were being measured across the circuit terminals $T$, according to Ohm's law, the resultant current would be 0.5 amp. This current is far in excess of the normal meter rating of 1 ma (0.001 amp), and the overload would undoubtedly seriously damage the instrument. Obviously, the internal resistance of the meter circuit must be increased to a minimum value that will keep the maximum current at the safe value of 1 ma. The necessary minimum resistance is easily computed as follows.

The lowest resistance in the external circuit that could be measured is, of course, zero ohms (short circuit). This condition could be simulated by short-circuiting the terminals $T$ in the diagram in Fig. 36(a). Then, by Ohm's law, with $E$ fixed at 1.5 v and $I$ fixed at 1 ma,

$$R = \frac{1.5}{0.001} = 1{,}500 \text{ ohms.} \qquad (61)$$

The necessary internal resistance of the circuit, therefore, must be 1,500 ohms, which includes *all* the internal circuit resistance; allowance must be made, however, for the resistance of the meter itself. If the instrument in Fig. 36(a) has a resistance of 30 ohms, it will be necessary to add a resistance of 1,470 ohms to complete the circuit properly. The complete circuit is shown in Fig. 36(b). It is common practice to make $R$ variable in order to compensate for the battery resistance as the battery deteriorates with age.

Ohmmeters are widely used as radio service instruments and wherever absolute precision is not of major importance. They are very flexible and easy to operate. The accuracy of a given ohmmeter is limited by the degree of accuracy with which the scale divisions can be read. For example, the ohmmeter illustrated in Fig. 36(b) will show a variation in current if a resistance as low as 1 ohm is connected across the terminals. The instrument, for this condition, will indicate slightly less than maximum current. Whether or not this difference in current can be accurately read will depend upon how large the meter scale is.

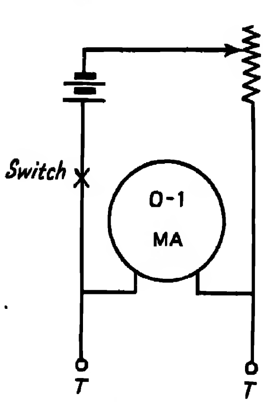


FIG. 37. Fundamental circuit for shunt type ohmmeter.

It is not feasible to construct a meter scale beyond a certain practical size. The range of an ohmmeter can be extended in either direction, however, very simply. If, for example, it is desired to *increase* the range of the instrument in Fig. 36(b) to read 10 times as much resistance at any given dial reading, the values of $E$ and circuit $R$ are increased 10 times. By Ohm's law then

$$I = \frac{E}{R} = \frac{15}{15{,}000} = 0.001 \text{ amp,} \qquad (62)$$

which is the proper current limit for the instrument used.

It is often convenient to *decrease* the resistance range of an ohmmeter. An ohmmeter having a range from zero to 300 ohms would have much more easily readable scale divisions. Thus, a resistance of $\frac{1}{2}$ ohm could easily be distinguished on such a meter. For this purpose, a shunt is inserted in the ohmmeter operating on the same principle as the shunt utilized in ammeter circuits. Figure 37 illustrates a typical **shunt-type** ohmmeter circuit. Observe that the resistance to be measured is shunted directly across the meter terminals. The principle of operation is clearly another application of the current rule for parallel circuits. With a very high value of unknown resistance or open circuit, very little or no current flows through the unknown external resistance. Therefore, a large part of the circuit current, or all of it, must flow through the meter circuit.

Conversely, with a very low value of unknown resistance, a large portion of the current flows through the external resistor and very little through the meter. Obviously, the ohms scale in a shunt-type meter is calibrated in the reverse direction from the series-type instrument, that is, maximum current through the meter indicates maximum resistance in the external circuit and vice versa. It should be noted that the meter is *always* in the circuit, so that, whether an external resistor is connected or not, current will flow through the meter. Therefore, a switch is provided to open the circuit when the instrument is not in use.

**Comparison Method.** The value of an unknown resistance may easily be determined by comparing it with a known value of resistance (see Fig. 38). $R$ is an accurately calibrated variable resistor, and $R_x$ is the unknown resistance to be measured. The switch is connected to the $R_x$ side of the circuit, and the current reading of the ammeter is noted. Switching to the $R$ side of the circuit, $R$ is varied until the ammeter indicates the same current reading. The two resistances are then equal in value, and the value of the unknown resistor can be read directly from the calibrated resistor $R$.

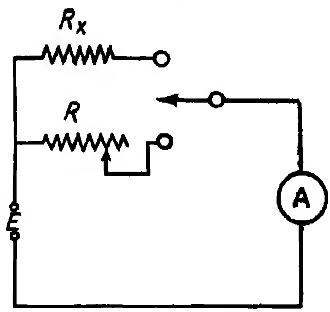


FIG. 38. Measuring resistance by the comparison method.

This method of measuring is usually restricted to laboratories, since it entails the use of accurately calibrated resistors with a high degree of precision.

**Wheatstone-Bridge Method.** The Wheatstone bridge far surpasses any other resistance measuring instrument in accuracy. It is a favorite laboratory instrument and is used wherever accuracy is of primary importance.

The voltage rule for series circuits states that the voltage across any part of a series circuit is directly proportional to the resistance of that part of the circuit. In Fig. 39(a), therefore,

$$\frac{E_1}{E_2} = \frac{R_1}{R_2}. \tag{63}$$

In Fig. 39(b) we have added another series leg of two resistors to the circuit, making a series parallel combination, and across the points $A$ and $B$ we have connected a very sensitive current meter. For the lower leg of this circuit, the series rule holds, and

$$\frac{E_3}{E_4} = \frac{R_3}{R_4}. \tag{64}$$

If the values of the resistors are such that no current flows through the

current meter, it is evident that no difference of potential exists between points $A$ and $B$. For this condition, we can say $E_1 = E_3$ and $E_2 = E_4$. It follows, then, that

$$\frac{E_1}{E_2} = \frac{E_3}{E_4}. \tag{65}$$

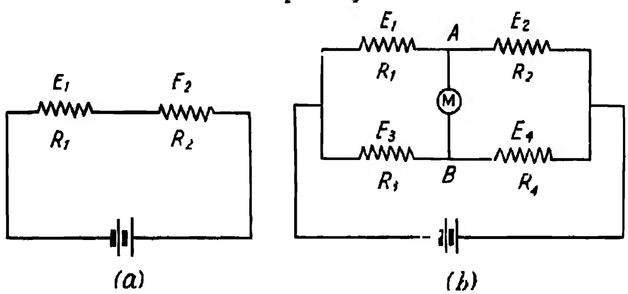



FIG. 39.

Substituting in Eq. (65) the equal values for these expressions given in Eqs. (63) and (64), we have

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}, \tag{66}$$

and, by the mathematical laws of ratio and proportion (Chap. II),

$$\frac{R_1}{R_3} = \frac{R_2}{R_4}. \tag{67}$$

If, in Fig. 39(b), the values of $R_1$ and $R_3$ are known and a calibrated variable resistance is inserted in place of $R_4$, we can readily ascertain the value of an unknown resistance inserted in place of $R_2$. This circuit is shown in the customary diamond-shaped diagram of the Wheatstone bridge in Fig. 40. Equation (67) is rearranged for this circuit as follows:

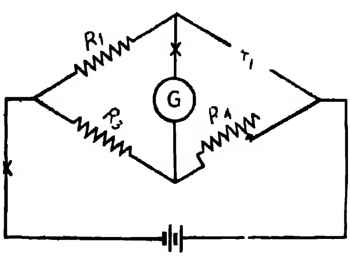



FIG. 40. Wheatstone bridge circuit

$$\frac{R_1}{R_3} = \frac{R_x}{R_4}. \tag{68}$$

The measurement is performed by inserting the unknown resistance across the $R_x$ terminals (Fig. 40). $R_4$ is then varied until no current flow is indicated by the meter. Zero current in this leg indicates that the bridge is balanced, that is, the voltages are in proportion, and Eq. (68) holds true. The fixed values of $R_1$ and $R_3$ and the value of $R_4$ as indicated by the calibration are substituted in Eq. (68), which is solved for $R_x$ as follows:

$$R_x = \frac{R_1 R_4}{R_3}. \tag{69}$$

Equation (69) is the conventional equation for the Wheatstone bridge. The meter used in this instrument is a very sensitive type of current meter known as a **galvanometer.** The normal zero-current position of a galvanometer needle is the center of the scale. The galvanometer is capable of indicating a flow of current in either direction.

## *VOLTAGE-DIVIDER CIRCUITS*

Almost all types of radio equipment require more than one value of d-c voltage for proper operation. A receiver, for example, may require

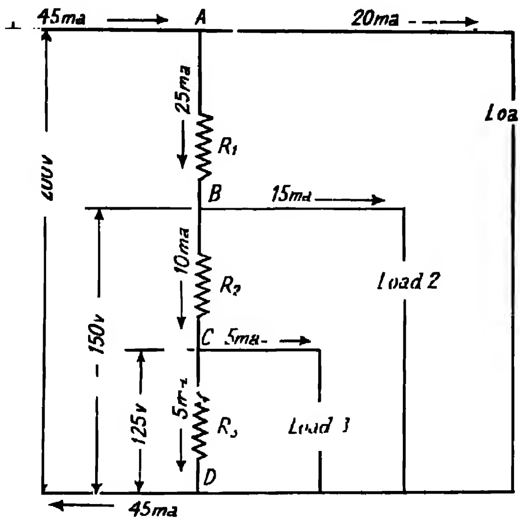



Fig. 41. A typical voltage-divider circuit.

150 v to supply the plate circuits of r-f tubes, 250 v for the output-stage plate circuit, and still other values for screen-grid and bias circuits. Naturally, it would be poor economy to incorporate separate voltage sources for each separate need in a given piece of apparatus. One power supply is usually adequate, except for certain special circuits.

The power supply for a typical receiver circuit is designed to deliver the highest voltage required by any of the receiver circuits. The lower voltages are procured from the same source by means of a special resistance network called a **voltage divider.** A diagram of a typical voltage-divider circuit is shown in Fig. 41. This particular voltage divider has been designed to supply power to three load circuits at different voltages: 20 ma at 200 v, 15 ma at 150 v, and 5 ma at 125 v.

The first step in designing such a circuit is to provide a source of power that will furnish the highest voltage and the total current required. The

highest voltage in this circuit is 200 v. We shall assume that our power supply is capable of delivering 200 v at a current drain of 45 ma. Note that an extra 5-ma current has been allowed. This extra current is to provide the "bleeder" current, about which more will be said later.

Since the power supply delivers 200 v, it can be directly connected across load 1. The next lower voltage, 150 v, is required by load 2. The necessary drop in voltage is obtained by inserting the resistor $R_1$ in the circuit, and the value of $R_1$ is computed by application of Kirchhoff's and Ohm's laws. At point $A$ according to Kirchhoff's law, there is as much current flowing from $A$ as there is flowing to it. We have 45 ma flowing *to* it from the power source; 20 ma is flowing *away* from $A$ through load 1. Therefore, the remaining 25 ma must flow *away* from point $A$ through the resistor $R_1$. This current value can be checked by adding the remaining load currents: 15 ma (load 2), 5 ma (load 3), and 5 ma (bleeder load), all of which must, of course, flow through $R_1$. Since the desired voltage drop across $R_1$ is 50 v, the required value of $R_1$ is

$$R_1 = \frac{E}{I} = \frac{50}{0.025} = 2{,}000 \text{ ohms}. \tag{70}$$

At point $B$, we have 25 ma of current flowing *to* point $B$ through $R_1$ and 15 ma flowing *away* from $B$ through load 2. Consequently, the remaining 10 ma must flow away from $B$ through $R_2$. $R_2$ is designed to drop the voltage from 150 v to 125 v for load 3, a difference of 25 v. Therefore,

$$R_2 = \frac{E}{I} = \frac{25}{0.01} = 2{,}500 \text{ ohms}. \tag{71}$$

At point $C$, we have 10 ma flowing *to* $C$ through $R_2$ and 5 ma flowing *away* from $C$ through load 3. The remaining 5 ma must flow away from $C$ through $R_3$. Since the remaining voltage differential between points $C$ and $D$ is 125 v, $R_3$ will be

$$R_3 = \frac{E}{I} = \frac{125}{0.005} = 25{,}000 \text{ ohms} \tag{72}$$

The question might reasonably be raised: Why is $R_3$ required in the circuit at all? The necessary voltage differentials have been obtained through resistors $R_1$ and $R_2$. Why waste 5 ma of current through $R_3$?

Actually, the 5 ma, or **leakage current,** flowing through $R_3$ serves a very useful purpose. If the **bleeder resistor,** as $R_3$ is called, were omitted from the circuit, it is true that the other two resistors would provide the necessary voltage drops very nicely, *provided the load remained essentially constant*. However, this condition is more often not true of radio equipment. Let us see what happens with no bleeder in the circuit when the load varies.

With $R_3$ out of the circuit, $R_1$ would carry 20 ma and $R_2$ would carry

5 ma. The necessary values of $R_1$ and $R_2$ would then be computed as follows:

$$R_1 = \frac{50}{0.02} = 2{,}500 \text{ ohms}, \tag{73}$$

and

$$R_2 = \frac{25}{0.005} = 5{,}000 \text{ ohms}. \tag{74}$$

Now suppose that the current drawn by load 3 is suddenly increased 1 ma and that the other loads continue to draw the same current. An additional current of 1 ma through the resistors $R_1$ and $R_2$ will result in an additional voltage drop of

$$0.001(2{,}500 + 5{,}000) = 7.5 \text{ v}, \tag{75}$$

thus reducing the voltage on load 3 to 117.5 v. Usually such a voltage drop would be very undesirable.

Similarly, a decrease in load current will cause an increase in voltage with comparable harmful effects.

Let us now consider the same condition with the bleeder resistor $R_3$ in the circuit. As before, the necessary resistor values for this case are $R_1 = 2{,}000$ ohms, $R_2 = 2{,}500$ ohms, and $R_3 = 25{,}000$ ohms. If the current drawn by load 3 increases 1 ma, the total current furnished by the power supply will not increase 1 ma because the current drawn by $R_3$ will decrease. If we designate the current through the resistor $R_3$ by $I_3$, that through $R_2$ by $I_2$, and so on, we may set up the relations

$$I_1R_1 + I_2R_2 + I_3R_3 = 200, \tag{76}$$

$$I_2 = I_3 + 0.006, \tag{77}$$

$$I_1 = I_2 + 0.015. \tag{78}$$

Substituting the value of $I_2$ derived in Eq. (77),

$$I_1 = I_3 + 0.006 + 0.015 = I_3 + 0.021. \tag{79}$$

Substituting in Eq. (76) the values of $I_2$ and $I_1$ derived in Eqs. (77) and (79), Eq. (76) becomes

$$(I_3 + 0.021)R_1 + (I_3 + 0.006)R_2 + I_3R_3 = 200. \tag{80}$$

Solving for $I_3$

$$I_3 = \frac{200 - (0.021)R_1 - (0.006)R_2}{R_1 + R_2 + R_3}, \tag{81}$$

or

$$I_3 = \frac{200 - (0.021)2{,}000 - (0.006)2{,}500}{2{,}000 + 2{,}500 + 25{,}000}, \tag{82}$$

and

$$I_3 = \frac{143}{29{,}500} = 0.00485 \text{ amp} = 4.85 \text{ ma}. \tag{83}$$

The voltage across load 3 is then

$$I_3R_3 = 121.2 \text{ v.} \tag{84}$$

This is a voltage drop of 3.8 v, or approximately half the drop which occurred when the bleeder was not in the circuit.

It can be seen, therefore, that a bleeder acts to stabilize the circuit voltage. Although in actual practice the bleeder does not *completely* stabilize the voltage fluctuation with variable loads as shown above, it very materially lessens the effects of such load variations. The larger the proportion of total current that flows through the bleeder, the better the regulation which is obtained. It is customary to design voltage-divider circuits so that the bleeder current is 10 per cent of the total current.

The power rating for resistors used in voltage dividers can easily be computed by applying the power equation to each resistor. The current through the resistor in each case should be multiplied by the voltage drop across it. Voltage dividers are often manufactured as a single resistor with taps at the proper points. With such a unit, the power dissipation of the portion consuming the most power should be used as a basis for the whole resistor. Of course, proper safety factor allowance should be made as described earlier in this chapter.

## *QUESTIONS AND PROBLEMS**

**1.** If the value of a resistance to which a constant emf is applied is halved, what will be the resultant proportional power dissipation?

**2.** State the three ordinary mathematical forms of Ohm's law.

**3.** What should be the minimum power-dissipation rating of a resistor of 20,000 ohms to be connected across a potential of 500 v?

**4.** If a relay is designed to operate properly from a 6-volt d-c source, and if the resistance of the winding is 120 ohms, what value of resistance should be connected in series with the winding if the relay is to be used with a 120-v d-c source?

**5.** A milliammeter with a full-scale deflection of 1 ma and a resistance of 25 ohms was used to measure an unknown current by shunting the meter with a 4-ohm resistor. It then read 0.4 ma. What was the unknown current value?

**6.** Given a milliammeter of full-scale deflection equal to 1 ma and an internal resistance of 50 ohms, what value of additional series resistance must be used to permit operation as a voltmeter with a full-scale deflection at 70 v?

**7.** Two resistances of 18 and 15 ohms are connected in parallel; in series with this combination is connected a 36-ohm resistance; in parallel with this total combination is connected a 22-ohm resistance. The total current flowing through the combination is 5 amp. What is the current value in the 15-ohm resistance?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

**8.** A 10,000-ohm 100-w resistor, a 40,000-ohm 50-w resistor, and a 5,000-ohm 10-w resistor are connected in parallel. What is the maximum value of total current through this parallel combination that will not exceed the wattage rating of any of the resistors?

**9.** Two resistors are connected in series. The current through these resistors is 3 amp. Resistance 1 has a value of 50 ohms, resistance 2 has a voltage drop of 50 v across its terminals. What is the total impressed emf?

**10.** Assume a resistance of 8 ohms in parallel with a resistance of 6 ohms; in series with this combination is a resistance of 77 ohms. What is the total resistance of this combination?

## Chapter V

# BASIC ALTERNATING-CURRENT THEORY

Thus far only simple unidirectional currents called "direct currents" have been discussed. The study of radio, however, is fundamentally the study of alternating currents, which is the technical name for currents that periodically change their direction in a circuit. Alternating currents are appreciably more useful than direct currents. It will be seen how this peculiar characteristic of periodic current reversals makes such currents infinitely more flexible to handle and, as a result, more adaptable for specific applications. Radio itself is but one of the many sciences that owe their inception to the application of alternating currents.

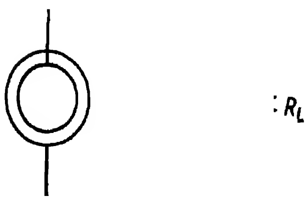



Fig. 42

Specifically, an **alternating current** *may be defined as a current that reverses its direction at regular intervals* (see Fig. 42). The alternator is connected to a load which is represented by the resistance $R_L$. If the alternating current supplied by the generator is of such a nature that it reverses its direction twice in each second of time, it is said that this alternating current has a *frequency of two cycles per second*. A *cycle* of alternating current refers to the period during which an entire round of changes in the current is completed, and the current returns to its original state. Immediately after returning to its original state the period commences to recur, that is, the second cycle begins.

Figure 43 is a graphical representation of the current in the circuit of Fig. 42. The passage of time is indicated from left to right on the $X$ axis. The amplitude of the current is indicated on the $Y$ axis, the $X$ axis being taken as the point of zero current. The amplitude of the current in one direction is indicated *above* the $X$ axis, the amplitude in the opposite direction is read *below* the $X$ axis.

The current rises from zero (point $A$) and continues to rise until the point of maximum amplitude (point $B$) is reached. During this process, however, a fraction of a second of time is consumed. For the two-cycle-per second current concerned, the time period is $\frac{1}{8}$ sec as indicated along the $X$ axis. During the second $\frac{1}{8}$ sec, the current, although still traveling in the same direction, *decreases* in amplitude until at the end of $\frac{1}{4}$ sec the amplitude is again at zero (point $C$). During the next $\frac{1}{8}$ sec, the current

rises from zero to maximum amplitude in the opposite direction (point *D*). The current then decreases in amplitude until at the end of ½ sec (point *E*) it is again at zero and the process is ready to recur. At this point (point *E*) an entire cycle has been completed.

Each cycle is composed of two *alternations*. An alternation is that portion of a cycle during which the current rises from zero, passes through maximum amplitude, and again drops to zero, traveling in *one direction*.

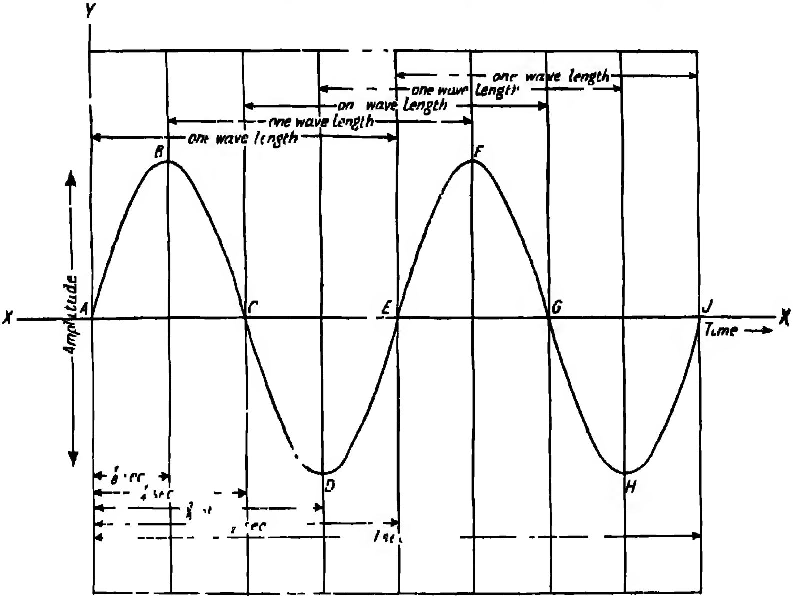



FIG. 43. Graph of an alternating current of two cycles per second.

Thus, in Fig. 43, curve *ABC* represents one alternation, curve *CDE* represents another, and so on. It follows, therefore, that there are two alternations to every cycle of alternating current.

The second is taken as the standard electrical unit of time. In the above illustration, since *two* cycles were completed in *one* second of time, the frequency of the current was said to be *two cycles per second*. If the length of time it takes to complete one cycle is 1/60 sec, as in commercial power circuits, the frequency is 60 cycles per second. Since the second is the standard unit of electrical time, the "per second" can be dispensed with, and the frequency of a given current can be expressed simply in cycles. Thus, the current in the circuit of Fig. 42 is called a "two-cycle alternating current." The "per second" is understood.

A person standing on a railroad depot platform could compute the length of each car in a train of freight cars passing the platform at a

constant rate of speed. This would simply be a matter of counting the number of cars which passed the platform in a given period of time. The length of each car could then be computed by dividing the speed of the train per unit time by the number of cars passing the platform in the same unit of time, that is, the *frequency* of their passing. Thus, if a train is moving at a constant speed of 1,000 ft per second and timing it with a watch reveals that 10 cars pass the platform in 1 sec, it follows that the length of each car is 1,000/10, or 100 ft.

Similarly, the length of one *wave* of alternating current can be computed. In Fig. 43, a wave length of alternating current is represented by the distance from $B$ to $F$, that is, from one crest to the next crest, traveling in the same direction. Of course, inasmuch as the actual distances are the same, a wave length can also arbitrarily be taken as the distance $AE$, $CG$, or $DH$. Since the velocity of an electric current is a constant, regardless of the circuit dimensions, the wave length can easily be computed as follows when the frequency is known.

$$\text{wave length} = \frac{300{,}000{,}000\ \text{m}}{\text{frequency}}, \tag{1}$$

where 300,000,000 m is the metric equivalent of 186,000 miles. The frequency must be expressed in cycles. The wave length obtained will be in meters.

Conversely, the frequency can be computed when the wave length is known, since Eq. (1) is a simple linear equation that can be solved for any unknown when the remaining two factors are known. Thus,

$$\text{frequency} = \frac{300{,}000{,}000\ \text{m}}{\text{wave length}}. \tag{2}$$

## *PRODUCTION OF AN ALTERNATING CURRENT*

**Generator Action.** In Chap. I, the laws governing the relation between magnetic and electric fields were discussed. These laws are

1. *A moving electric field creates a magnetic field.*
2. *A moving magnetic field creates an electric field.*

It should be noted that a magnetic field must be *moving* in order to create an electric field. Thus, a conductor through which is flowing a direct current is surrounded by a magnetic field created by this current flow or *moving electric field*. Since the current, being *direct* current, is constant in amplitude, however, the magnetic field created by it does *not* move. As a result, there is no electric field created by such a magnetic field; that is, if a second conductor were placed in this magnetic field, no current would flow in this conductor, because there is no electric field present. At the instant the circuit of the original conductor was closed,

the current started rising from zero to maximum amplitude. During the small fraction of time it took to do this, the magnetic field was in motion, as it expanded from nothing to its maximum pattern. Therefore, an electric field was created by the *moving* magnetic field during this time. Once the current has reached maximum amplitude, however, there is no longer any variation in amplitude, no resultant *moving* magnetic field, and, consequently, no electric field created. Similarly, when the current is cut off from such a circuit, the current drops from maximum value to zero, and the magnetic field collapses. In doing so, the resultant motion of the magnetic field again creates an electric field. It can thus be seen that a *moving* magnetic field results only when an electric field varies in amplitude. One of the principal advantages of an alternating current over a direct current is that the alternating current is constantly varying in amplitude, thereby constantly creating a moving magnetic field. Any conductor placed within the field of a conductor bearing alternating current will have a flow of current (a moving electric field) induced in it. This accounts for the wide adaptability of alternating currents in any application utilizing the principle of induction.

Direct currents are often made to create moving magnetic fields by rapidly interrupting the circuit artificially, causing the magnetic field alternately to expand to maximum and to collapse to zero and thus creating an electric field. Such *pulsating direct currents* are utilized for applications where alternating currents are not readily available, as in automobile radio equipment. In the latter application, a vibrator is used to interrupt the direct current obtained from the automobile battery. The principle of operation is similar to the old time spark coil.

The laws governing the relation of magnetic and electric fields received their first basic applications in generators and these laws are often referred to as the two basic laws governing generator action.

To induce an electric voltage in a conductor, the primary requisite is that the magnetic field in which it is placed be *moving* with respect to the conductor, and in such a direction that the conductor cuts the magnetic lines of force. It has been determined experimentally that the same result is obtained if the conductor is moved through the magnetic field instead of having the magnetic field moved through the conductor. This principle is utilized in the electric generator, where a powerful magnetic field is maintained at a constant amplitude between two opposite magnetic poles. These are called the "field poles" of the generator. A conductor, wound about a high permeability iron core, called the "armature," is rotated by external mechanical means through the region of the magnetic field between the pole pieces. An elementary two pole generator of this type and a graph of the voltage produced by it are shown in Fig. 44.

Consider the movement of a portion of the conductor rotating about

the axis $O$ from point $A$ to point $B$ in Fig. 44. At point $A$, the direction of motion at that instant is *along* the lines of force, and, hence, no lines of force are cut; that is, zero voltage is induced in the conductor. At point $B$, the conductor is at the point nearest the north pole piece where the rate of cutting of magnetic lines is greatest; hence, the resultant induced voltage is maximum at this point. This is represented on the graph as the voltage value at point $B'$. As the portion of conductor continues its rotation from point $B$ to point $C$, it again passes from the position of maximum cutting of the field to no cutting, and the resultant

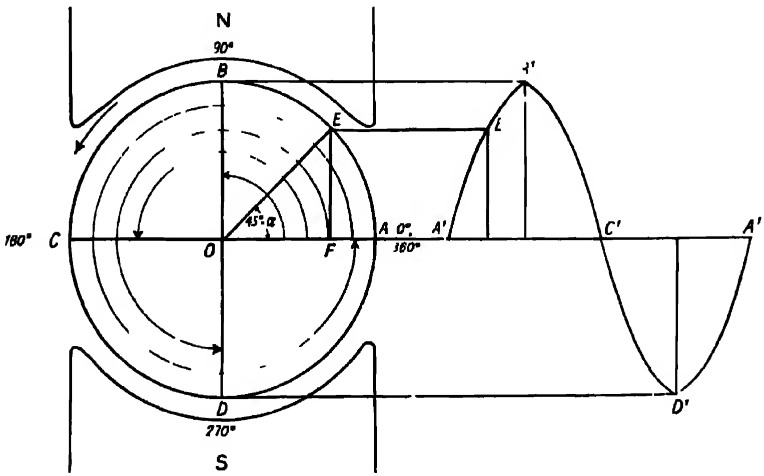



FIG. 44. Projection of an a-c sine wave.

induced voltage again falls to zero as indicated by point $C'$ on the graph. Moving from $C$ to $D$, the conductor again moves to a position where it is cutting a maximum number of lines of force per second and the voltage is again at a maximum. Here, however, the conductor moves through the field in the opposite direction. The resultant induced voltage, therefore, is of opposite polarity, and the current that flows as a result of this voltage flows in a direction opposite to that which was caused to flow during the conductor rotation from $A$ to $B$ to $C$. Finally, completing its revolution the conductor moves from point $D$ to point $A$, its original starting position, and the resultant induced voltage falls off to zero.

The conductor has now completed its movement through the 360° of a circle that comprise one revolution. In so doing, it has generated one complete cycle of alternating voltage. As a result of the induced voltage, an alternating current has been caused to flow, which varied in direction as the voltage varied in polarity. The cycle of alternating voltage, or current, is arbitrarily subdivided into 360 parts corresponding to the

360° of rotation that produced it. For simplicity, the designation "degrees" is maintained. This designation, which is called **phase angle,** provides a convenient method of referring to the instantaneous value of an alternating current at any point throughout the cycle.

The maximum value of current generated is represented graphically by the distance the point $B'$ is above the zero line. This distance is a projection of the line $OB$ from the generator, since $OB$ represents the position of the conductor at the instant of maximum amplitude. The amplitude at any other instant will be some portion of the maximum value represented by $OB$, the exact value depending upon the position of the conductor at that instant. Thus, at point $E$, or 45°, the amplitude of the current will be proportional to the value of the line $EF$. The current value, therefore, in terms of the maximum current, will equal $EF/OB$ times the maximum. $OB$, however, is the radius of the circle. Since the radius of a given circle is the same at all parts of the circle, $OB$ equals $OE$, and the amplitude at 45° will equal the maximum current times $EF/OE$. In the triangle $OEF$, $EF/OE$ is the trigonometric sine of the angle $\sigma$ of 45°. Therefore, the current or voltage value at the phase angle of 45° is sin 45° times maximum or peak value. Thus, if the peak value or voltage in a given circuit is 155 v, the value at a phase angle of 45° is sin 45° times 155 = 0.707 · 155 = 110 v. The instantaneous value of voltage or current at any part of the cycle (any phase angle) can, therefore, be found by multiplying the maximum value by the sine of the phase angle.

Since the graph of Fig. 44 represents an infinite number of consecutive instantaneous values of amplitude, it is actually a projection of the sines of all the angles from 0 to 360°. For this reason, it is called a **sine curve,** or **sinusoid,** and all alternating currents that vary in this symmetrical manner are called "sinusoidal currents."

## *QUANTITATIVE VALUES OF ALTERNATING CURRENT*

An alternating current is continually varying in amplitude. The instantaneous value, or the value at any particular instant, can be computed as shown above. Which of these many instantaneous values should be used when referring to a given alternating current in a quantitative sense? It is obvious that some value must be chosen that is representative of the actual *work* done by the alternating current. Such a representative value is necessary in order to apply to alternating currents the laws for electric circuits such as Ohm's law, Kirchhoff's laws, the power equations, and so on. Since these laws already hold true for direct currents, the logical solution is to compare the work done by a given unit of direct current with the same amount of work done by a certain value of alternating current. The simplest way to accomplish

this is to compare the *heating effects* of the two types of currents, since this is one of the few forms of electrical work that is independent of the direction of the flow.

In a d-c circuit the power dissipated in the form of heat is equal to $I^2R$ according to the power equation. Thus, if a current of 10 amp flows through a circuit having a resistance of 5 ohms, the power dissipated in the form of heat in this circuit will equal $10^2 \cdot 5$, or 500 w. If the current value in this circuit is doubled, the resultant power dissipation will be $20^2 \cdot 5$, or 2,000 w. It is apparent then, that *the heating effect of a current is proportional to the square of the current.* In the above illustration, doubling the current produced four times as much heat. Similarly, tripling the current (increasing it to 30 amp) would produce nine times as much heat, represented by 4,500 w of power dissipated in the circuit, and so on. This relation of current to power consumed (heat dissipated) is known as the **current-square law.**

**Average Values.** Offhand, it would perhaps appear that a truly representative value of alternating current could be obtained by striking an average of the instantaneous values throughout a cycle. Such an average value is obtained by computing a number of consecutive instantaneous values of a current whose peak, or maximum value, is known and then averaging them. Thus, in Fig. 45 the 12 instantaneous values at 30°, 60°, 90°, 120°, and so on, are computed and averaged for an alternating current having a peak value of 100 amp. The instantaneous values are computed as shown above by multiplying the sine of the phase angle in each case by the peak value of 100 amp. The sum of the instantaneous values thus found is divided by their number, 12, to obtain the average value of approximately 63.6 amp. It follows that the greater the number of instantaneous values taken for this calculation, the more accurate the result. By dividing 63.6 amp by the peak value of 100 amp, a constant (0.636) is obtained that gives the specific relation between peak and average values for all cases. Expressed as formulas,

$$\textit{average}\ \text{value} = 0.636 \times \textit{peak}\ \text{value}; \tag{3}$$

$$\text{average} \begin{Bmatrix} \text{voltage} \\ \text{or} \\ \text{current} \end{Bmatrix} = 0.636 \times \text{peak} \begin{Bmatrix} \text{voltage} \\ \text{or} \\ \text{current} \end{Bmatrix}. \tag{4}$$

It has been determined, however, that the average value of an alternating current thus obtained is *not* an accurate expression of the effectiveness of the current. In other words, the 63.6 amp (average value) of alternating current does not have the same heating effect as 63.6 amp of direct current, that is, it does not accomplish the same amount of work. The average value, therefore, cannot be taken as representative of the effectiveness of an alternating current. Nevertheless, the method of

computing average values of alternating current as outlined in Eqs. (3) and (4) should be remembered, since average values are very useful for specific applications to be discussed later in the text.

**Effective Values.** A truly representative value of alternating current is that value which has the same heating effect as an equal value of direct current. Therefore, this value is directly derived from the current-square law, since this law governs the relation between current value and resultant heating effect. According to this law, the heating effect varies as the *square* of the current. An average of the *squares* of the instantaneous

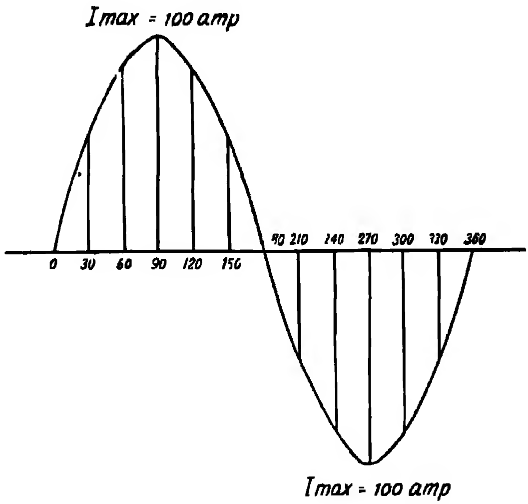



Fig. 45 Instantaneous values of alternating current

current values is therefore taken. The instantaneous values are computed as formerly. They are then *squared*. The sum of these squares is then divided by the number of values taken. The square root of this sum is the value of the current thus represented and is called the **root mean square** value, since it is the square *root* of the *mean* (average) of the *squares* of the instantaneous values. Inasmuch as this value is an accurate representation of the *effectiveness* of an alternating current in accomplishing work, it is referred to as the **effective value.** The effective value thus obtained for an alternating current has been found to be 0.707 times the peak value. Expressed as a formula

$$\textit{effective}\ \text{value} = 0.707 \times \textit{peak}\ \text{value}. \tag{5}$$

Conversely, by reversing the calculation,

$$\textit{peak}\ \text{value} = 1.414 \times \textit{effective}\ \text{value}, \tag{6}$$

Or

$$effective \begin{Bmatrix} \text{voltage} \\ \text{or} \\ \text{current} \end{Bmatrix} = 0.707 \times peak \begin{Bmatrix} \text{voltage} \\ \text{or} \\ \text{current} \end{Bmatrix}, \tag{7}$$

and

$$peak \begin{Bmatrix} \text{voltage} \\ \text{or} \\ \text{current} \end{Bmatrix} = 1.414 \quad effective \begin{Bmatrix} \text{voltage} \\ \text{or} \\ \text{current} \end{Bmatrix}. \tag{8}$$

The effective value is the value indicated on all voltmeters and ammeters. In all applications of alternating currents, values used are effective values unless specifically stated otherwise. Thus, when it is said that the ordinary house-lighting circuit operates at 110 v, that figure is the effective value of the voltage. The peak voltage for this circuit is 110 × 1.414, or approximately 155 v.

It should be noted that components utilized in a c circuits should be rated to withstand safely the *peak* voltages of the circuit. It is standard practice to add a generous safety factor to the peak voltage rating to allow for transient surges of voltage. Thus, in the house-lighting circuit mentioned above, the wire insulation must be able to withstand safely the peak voltage of 155 v in addition to possible transient voltages considerably higher.

**Frequency.** Certain characteristics of alternating currents are known to vary with the frequency. Consequently, alternating currents have been classified in four broad frequency groups: *power* frequencies, *audio* frequencies, *radio* frequencies, and *ultrahigh radio* frequencies There is no sharp dividing line between these groups that is there is no border line frequency at which an alternating current whose frequency is being changed suddenly loses the characteristics of one group and acquires those of the next group.

**Power frequencies** are the frequencies that have been found most efficient for the transmission of power Generally speaking, power frequencies are very low frequencies. Typical power frequencies in common use are 25 c and 60-c frequencies.

As the frequency of an alternating current is increased above the power frequencies, it enters into the range known as **audio frequencies.** The human ear is sensitive to sound waves which vibrate from 20 or 30 to 16,000 times per second. Such sound waves can be created by any physical vibration occurring within this range, such as the vibration of the human vocal cords. An alternating current can be made to cause a physical vibration by passing it through certain types of apparatus, such as an earphone. In so doing, the current causes a diaphragm (a small, thin metal plate) within the earphone to vibrate physically and set up sound waves. The frequency of this vibration is exactly equal to

the frequency of the alternating current causing it. If these vibrations fall within the range of the human ear (20 to 16,000 times per second), the frequency of the alternating current creating them is called an **audio frequency,** since it is capable of setting up audible sound waves under these conditions. Any alternating current, therefore, having a frequency which falls in the range of 20 to 16 000 c can be called a current of audio frequency. The theory and application of sound-conversion instruments, such as the earphone are taken up in detail in Chap. XIV.

As the frequency of an alternating current is increased beyond the range of audio frequencies it enters that portion of the frequency spectrum known as **radio frequencies.** Radio frequencies are those at which an alternating current possesses the property of radiating a magnetic field of appreciable energy out into space at the velocity of light. This phenomenon is the basis of the science of radio. Since an entire chapter is devoted to the theory of wave propagation in this book, the subject will not be enlarged upon at this time. Suffice it to say that radio frequencies extend from the higher portion of the a-f band to somewhere in the vicinity of light frequencies.

The radio frequencies above 30,000,000 c (30 megacycles) are known as **ultrahigh radio frequencies** or **ultrahigh frequencies.** As the frequency is increased above 30 megacycles, the radiated magnetic field begins to take on more and more of the properties of light. Ultrahigh frequencies cannot be transmitted beyond the line of sight, as can other radio frequencies, for example; nor can they be transmitted through opaque substances as can the lower frequencies. This similarity to light frequencies is the primary distinction between ultrahigh and radio frequencies

## *A-C METERS*

The instruments used in measuring a-c values differ in several respects from those used in d-c measurements. The moving-coil type of meter used for d-c measurements cannot be used for a-c measurements. During one alternation of the cycle, current would flow through such a meter in one direction with a resultant pointer deflection, on the following alternation, the current would flow through the instrument in the opposite direction and the pointer would deflect in the opposition direction. Actually, since the alternations of even an l-f alternating current follow each other with extreme rapidity, the net result would be no movement of the pointer at all. The natural inertia of the meter pointer would prevent it from even getting started in one direction before the following alternation tended to draw it in the opposite direction. The extent of the total pointer movement of such an instrument would be a slight quivering about the zero mark.

**The A-C Ammeter.** There are three main types of a-c ammeters: the

**rectifier** type, the dynamometer type, and the movable-iron type. The **rectifier-type** instrument is a movable-coil permanent-magnet current meter which has been adapted for use with alternating current by the insertion of a rectifier in the meter circuit. A rectifier is a device which

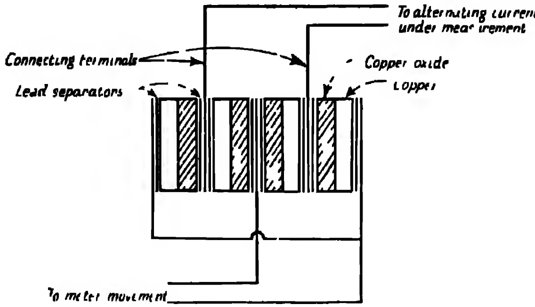



FIG. 46. Full-wave copper oxide rectifier.

converts alternating current into pulsating direct current. Rectifiers in general use are discussed at length in Chap. XII.

The rectifiers used in meter circuits are of the copper oxide type. Despite the fact that copper itself is a very good conductor the oxide

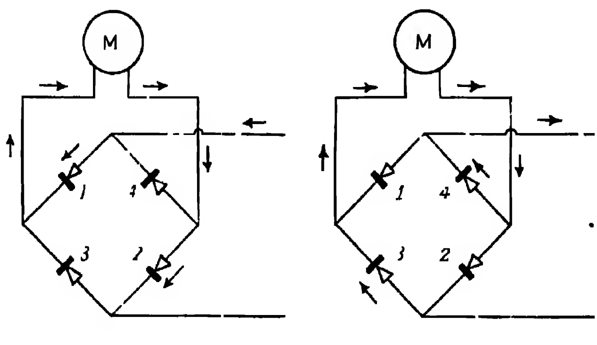



FIG. 47. Schematic diagram of full-wave copper oxide rectifier.

of copper was found to possess the remarkable characteristic of permitting the flow of current through it in *one direction only.* Figure 46 illustrates a typical copper oxide rectifier for meters. Commercial copper oxide rectifiers are composed of a series of copper plates that have been oxidized on one side and are connected by alternate lead plates. The circuit of a full-wave meter rectifier is shown in Fig. 47. On the first alternation, the current is allowed to pass by rectifier units 1 and 2 –in Fig. 47(a) but is rejected by units 3 and 4. The current movement through the meter is shown from left to right. On the following alternation, the current is rejected by units 1 and 2 Fig. 47(b)—but is passed

by units 3 and 4. The current movement through the meter is still from left to right.

Figure 48 illustrates an alternating current and the resultant pulsating direct current after it has passed through a circuit such as that of Fig. 47.

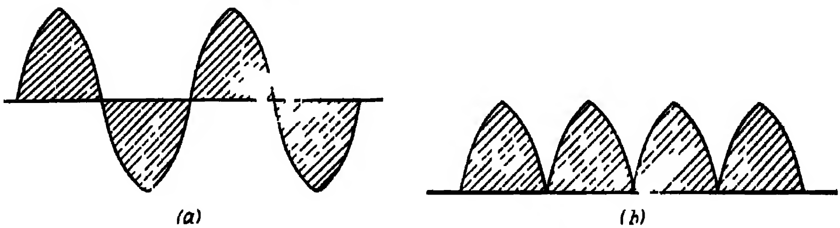



FIG. 48 Input and output currents in full ... ctifier (a) Alternating-current input (b) Pulsating direct current output

It will be noted that although the current in Fig. 48(b) is traveling in one direction only, it is by no means constant in amplitude. The peaks and wave shapes are still identical with the original alternating current, but there are no reversals. When such a pulsating direct current is passed

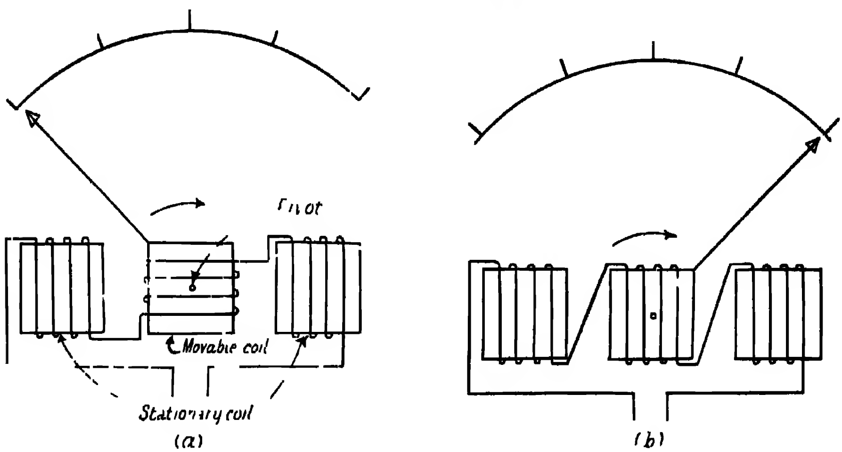



FIG. 49 Elementary dynamometer-type ammeter movement (a) Zero deflection. (b) Full scale deflection

through a meter, the current value indicated by the meter will be the *average* value of the alternating current. In other words, the meter reading will be 0.636 times the peak value. Since the value desired is the *effective* value, the meter value must be multiplied by 0.707/0.636 or 1.1 to obtain the effective value.

In factory-constructed rectifier-type meters this correction has already been made on the dial calibration, and the effective value may be read directly from the meter. If a regular d-c meter is made to serve as an a-c

instrument by adding a rectifier to the circuit, it should be remembered that the dial value as read on the meter must be multiplied by the correction factor 1.1 to obtain effective value.

The **dynamometer** type of ammeter depends for its operation upon the repulsion produced between a stationary coil and a movable coil by the action of their magnetic fields. A diagram of an elementary dynamometer-type ammeter is shown in Fig. 49. The movable coil is mounted on a pivot within the stationary coil, and the windings of the two coils are connected in series. When the movable coil is at zero position, as illustrated in Fig. 49(a), and a current is passed through the coils the magnetic fields thus created cause repellent and attractive forces of such a nature as to rotate the movable coil about its axis in clockwise direction. The resultant movement to full-scale deflection is shown in Fig. 49(b). The reversals of alternating current do not affect the operation of this instrument since a reversal of current causes a reversal of the magnetic fields of *both* coils. Consequently, the repellent and attractive forces and the meter movement continue in the same direction. The dynamometer type ammeter indicates the effective or root mean square value of alternating current. In common with all ammeters, this instrument is connected in series with the circuit to be measured.

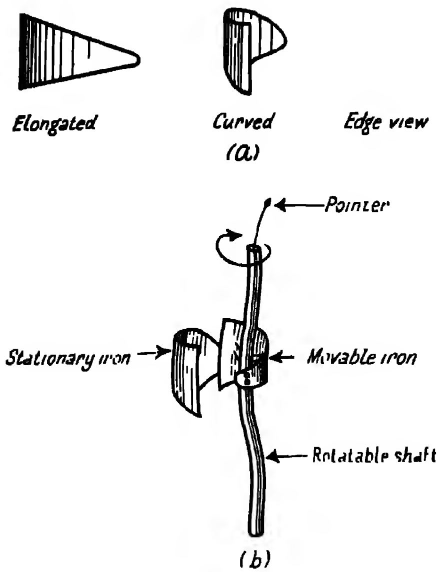



FIG. 50. Components of movable movement (a) Construction of iron vane (b) Vanes concentrically mounted

The **movable-iron** type of ammeter depends for its operation upon a principle similar to that of the dynamometer instrument. In this instrument, a small thin triangular plate of soft iron is bent in the shape of a half cylinder, as shown in Fig. 50(a), and is secured to a vertical shaft so supported that it can turn freely on jewel bearings. Since this is the movable part of the instrument, it is called the **meter armature.** Another wedge-shaped, thin iron plate is bent in the shape of a half cylinder and mounted concentrically about the armature. (See Fig. 50(b).) This latter iron is stationary. This arrangement permits the armature to rotate about its axis within the curvature of the stationary iron. The entire mechanism is mounted within a coil, as shown in Fig. 51.

When current flows through the coil on one alternation, both iron pieces are magnetized and assume north and south poles. Since the irons are *parallel*, the upper edges of both pieces and the lower edges of both pieces will have like polarities. Thus, at a given instant, both upper edges could have a north polarity and both lower edges a south polarity.

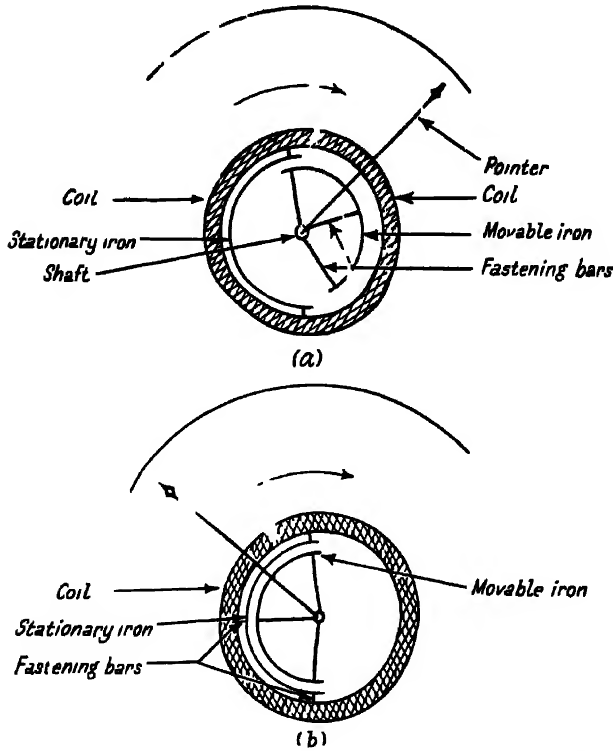



FIG. 51 Arrangement of fixed and movable iron vanes within coil. Note the fastening bars with which the stationary iron is attached to the coil form and by means of which the movable vane is mounted on the pointer shaft. (a) Full scale deflection. (b) Zero deflection.

Since like poles repel each other, the magnetic fields of both irons would be of such a nature as to cause mutual repulsion between the iron pieces at all points. Since the stationary iron is wedge shaped, however, the magnetic field will be stronger about the butt or longer side. The resultant greater repulsion at this end of the iron causes a movement of the armature toward the shorter pointed end of the stationary iron. Because the strength of the magnetic field is a function of the current intensity, the movement of the meter armature is a direct indication of the amplitude of the current in the coil. A pointer attached to the armature shaft

moves over a calibrated dial as the shaft rotates, and the current value is read directly from this dial.

On the following alternation, the current flow through the coil is in the opposite direction. The irons within the coil are magnetized in the opposite direction. Since the magnetic fields of *both* irons are reversed, however, the mutual repulsion of like poles continues, and the pointer movement is in the same direction as for the first alternation. The movable-iron type of ammeter indicates effective values of current.

The force on the dynamometer-type ammeter is proportional to the product of the current strength in one coil and the magnetic field strength of the other coil. The force on the movable-iron-type ammeter is proportional to the product of one magnetic pole strength and the magnetic field from the other pole. In either case, the force and, hence, the pointer movement are proportional to the *square* of the current. Assume, for example, that an effective current of 1 amp in a given meter results in a circular pointer movement of 5°. A current of 2 amp through this instrument will cause a pointer movement of 20°. A current of 3 amp will cause a pointer movement of 45°, and so on. The calibrated scale on these instruments, consequently, is nonuniform, and the scale divisions are more crowded on the lower amplitude end of the scale. A typical a-c ammeter scale is illustrated in Fig. 52. Alternating-current ammeters of this type are often called **current-square meters** because of the above reasons.

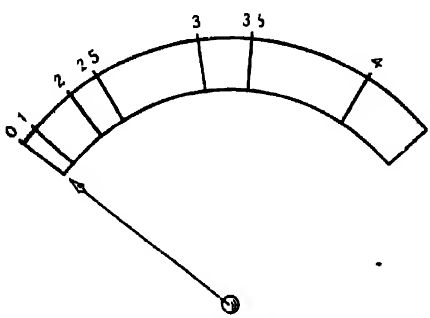



Fig. 52 Scale calibration of a typical current square meter.

According to the basic current law for series circuits, the current is the same in all parts of a series circuit. For that reason ammeters are always connected in series with the circuit whose current is being measured. In order to minimize the disturbance that the internal resistance of a meter might create in the circuit, the resistance of ammeters is always kept as low as possible. Consequently ammeter coils are constructed of comparatively few turns of wire of relatively large gage.

**Special-Purpose Ammeters.** At the extremely high frequencies encountered in radio work, it was found that the inherent inductance of the moving-coil types of ammeters introduced serious changes in a circuit. Such instruments therefore, cannot be used to measure accurately currents at radio frequencies. A different type of ammeter operating on the thermal principle was designed to fulfill this need. Thermal ammeters are of two types, namely, the hot-wire ammeter and the thermocouple ammeter.

The **hot-wire ammeter** depends for its operation upon the physical expansion that occurs in a very fine platinum-silver wire when current is passed through it.

A diagram of an elementary hot-wire ammeter is shown in Fig. 53. As current passes through the wire *AB*, the heat generated causes the wire to expand. A silk thread *S* is attached to the platinum-silver wire and wound around the drum *D* with its other end terminating in a spring. When current flowing through the wire causes it to expand, the spring acting on the silk thread takes up the slack, and the movement of the thread *S* causes the drum *D* to revolve in a clockwise fashion. A pointer

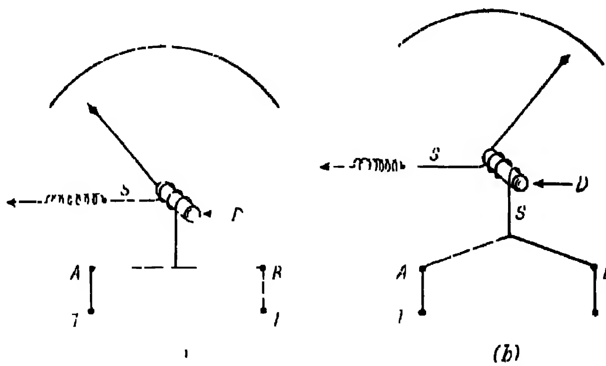



FIG. 53. Principle of the hot-wire ammeter. (a) Zero current, no deflection. (b) Maximum current, wire expanded, full-scale deflection.

secured to this revolving drum moves over a graduated scale calibrated in amperes. When current flow is discontinued, the normal contraction of the wire as it cools causes the pointer to move back to zero position.

Since only a very small piece of wire is included in the circuit to be measured, the inductive effect of this instrument is practically negligible. Nevertheless, the hot-wire ammeter has several disadvantages. The fragility of the hot wire does not permit the instrument to be subjected to very rugged use. The thermal expansion characteristic of the platinum-silver wire causes a lag in the operation of the instrument. Several seconds elapse after a current is passed through the instrument before the pointer registers accurately. In addition, the instrument is affected by room temperatures and zero adjustment must be corrected often. For most applications, the hot-wire ammeter has been superseded by the thermocouple ammeter.

The **thermocouple ammeter** depends for its operation upon the emf produced by heating dissimilar metals. It has been found that steel and constantan, bismuth and antimony, and certain other pairs of dissimilar metals will produce a d-c emf when brought into contact and subjected

to the effects of heat. Figure 54 illustrates a thermocouple-ammeter circuit. Two dissimilar metals $AC$ and $BC$ are brought into contact with a conducting wire at point $C$. When current flows through the wire $DE$, the heat generated at the point $C$ causes an emf to be generated by the action of the dissimilar metals $AC$ and $BC$. This emf is applied to the terminal of a sensitive millivoltmeter of the permanent-magnet-moving-coil d-c type. The latter instrument operates in its usual manner when this emf is applied to it. The scale of the instrument, of course, is graduated in amperes.

The thermocouple ammeter combines the advantages of the hot wire with the accuracy of the d-c voltmeter.

Both thermocouple and hot-wire ammeters are current square meters, since both depend for their operation upon the heating effects of current. Both instruments indicate effective values and may be used for either direct or alternating current. Because the inductance and capacitance of these instruments are negligible, their operation is independent of frequency.

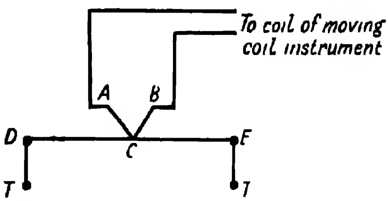


FIG. 54. Principle of the thermocouple.

**The A-C Voltmeter.** The a-c voltmeters are classified in the same way as the ammeters, that is, rectifier, dynamometer, and movable iron types. The rectifier-type a-c ammeter is converted to voltmeter use by the same method as its d-c prototype; that is, a suitable multiplier resistor is inserted in the circuit, and the calibration is changed to read volts instead of amperes. Since the principle is identical with that of the d-c voltmeter it need not be reviewed here.

Actually, all instruments whether voltmeters or ammeters, are *current measuring* instruments. According to the Ohm's law equation $E = IR$, if the resistance of a circuit is kept constant the voltage varies directly as the current. Since the resistance of any given measuring instrument is a constant for that instrument regardless of the use to which it is put the current through such an instrument is directly proportional to the voltage across it. Since a voltmeter must be connected directly across the voltage source to be measured, that is, in parallel one of the primary prerequisites of a voltmeter is that it be of high internal resistance. A low-resistance voltmeter connected directly across a voltage source will draw excessive current from the circuit, thereby diverting current from its normal course through the load circuit of the network under consideration. Therefore, a voltage reading obtained with a low-resistance meter will not be a true indication of the voltage in such a circuit under normal load conditions.

When a permanent-magnet instrument (such as the d-c or rectifier

type) is utilized as a voltmeter, the resistance is increased by the insertion of an external *multiplier* resistor in series with the meter internal resistance. This principle was discussed in an earlier chapter. Since the magnetic field produced by the dynamometer- or movable iron-type instruments is located practically all in air, it is relatively weak, and such instruments are generally less sensitive than the permanent magnet type.

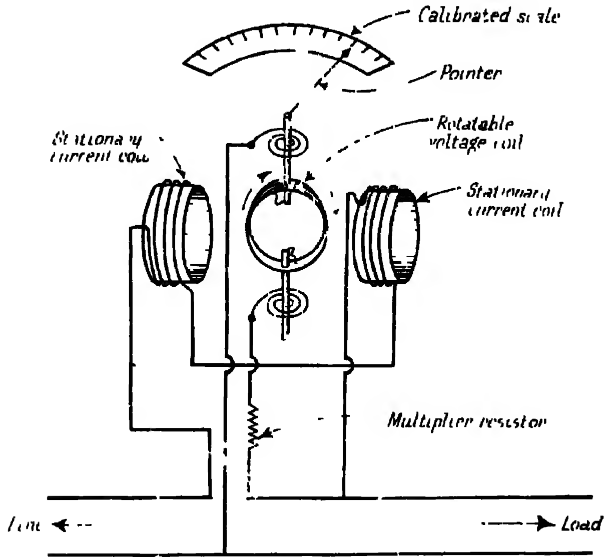



FIG. 55. Elementary wattmeter.

If an instrument of the dynamometer or movable iron type were utilized with a multiplier resistor to measure voltage, the small current resulting from the high meter circuit resistance would be insufficient to actuate the instrument. Consequently, the resistance of such instruments is augmented by increasing the number of turns in the meter coil instead of by adding external resistance. Since the strength of a magnetic field is a function of the product of the current times the number of turns in the coil, the loss of magnetic field caused by the smaller current flow is offset by the increase in magnetic field caused by a greater number of turns.

In general, therefore it may be said that voltmeters and ammeters of the dynamometer or movable iron type are identical in construction except for the fact that voltmeter coils contain many hundreds of turns of wire, whereas ammeter coils contain comparatively few turns.

**The A-C Wattmeter.** The type of wattmeter in general use operates on the principle of the dynamometer. The circuit of a typical elementary wattmeter is shown in Fig. 55. Since the power consumed in a circuit is

the product of the voltage and amperage, the circuit combines the voltmeter and ammeter connections. Inasmuch as the dynamometer principle has already been discussed, the circuit of Fig. 55 should be self-explanatory.

Factually, the power factor must be considered when computing power in an a-c circuit. The wattmeter automatically takes into account the power factor, which will be discussed in Chap. IX

### *QUESTIONS AND PROBLEMS**

**1.** What single instrument may be used to measure electrical resistance? Electric power? Electric current? Electromotive force?

**2.** What is the ratio of peak to effective voltage values of a sine wave?

**3.** If a d-c voltmeter is used to measure effective alternating voltages by the use of a bridge-type full-wave rectifier of negligible resistance, by what factor must the meter readings be multiplied in order to give corrected readings?

**4.** By what factor must the voltage of an a-c circuit as indicated on the scale of an a-c voltmeter, be multiplied in order to obtain the average voltage value?

**5.** What type of meters may be used to measure r-f currents?

**6.** A voltmeter is described as having "1,000 ohms per volt." What current is required to produce full-scale deflection?

**7.** If two voltmeters are connected in series, how would you be able to determine the total drop across both instruments?

**8.** If two ammeters are connected in parallel, how may the total current through the two meters be determined?

**9.** Is the angular scale deflection of an ammeter of movable iron type proportional to the square or the square root of the current, or merely directly proportional to the current?

**10.** Describe the construction and characteristics of a dynamometer-type indicating instrument.

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

## Chapter VI

# MOTORS AND GENERATORS

The study of motors and generators is an important phase in the treatment of radio communication. Practically every radio station, whether afloat, aloft, or ashore, utilizes motors or generators, or both, for some important application. The technician charged with the maintenance of a radio station, therefore, must have a working knowledge of these machines. Such a working knowledge can only be predicated upon a thorough understanding of the fundamental principles of operation involved. No man can hope successfully to repair a disabled motor or generator unless he first understands how it is supposed to operate. In addition, if the underlying theory of operation is understood, the tendency to abuse or overload a machine is eliminated. Breakdowns are consequently less frequent with a resultant decrease in maintenance overhead.

Motors and generators are best studied separately. Generators may be classified in two broad divisions: a-c generators and d-c generators. Each classification may be further subdivided: a-c generators are divided into three general types, namely, the revolving-field alternator, the revolving-armature alternator, and the inductor type alternator; and d-c generators are also divided into three general types, namely, the series-wound dynamo, the shunt-wound dynamo, and the compound-wound dynamo.

Similarly, motors can also be classified in two broad divisions—a-c motors and d-c motors. Each of these classifications is further subdivided. There are several general types of a-c motors, but the one type in general use today is the repulsion induction motor. Therefore, this is the only type of a-c motor that will be discussed in this chapter. Direct-current motors are divided into three types: the series-wound motor, the compound-wound motor, and the differential compound-wound motor.

In addition to the various types of motors and generators outlined above, there are a number of machines often used in radio stations that may loosely be called combinations of motors and generators. These include motor generators, dynamotors, and rotary converters. Each of these machines is discussed in a separate section of this chapter.

### *THE A-C GENERATOR*

**Fundamental Theory of Generator Action.** The basic theory underlying the generation of an alternating current was discussed in Chap. V.

The student is urged to reread this section (page 98) before continuing with the present discussion. The electric current producted by *any* type of generator discussed here is *alternating* current. The basic difference between d-c and a-c generators lies in the method of taking the current from the machine. In d-c machines, the current produced in the windings of the armature is alternating current. It is converted to direct current in the process of being collected, or transferred from the armature to an external circuit. This process is called **commutation.** Commutation is therefore only employed in d-c generators. It is accomplished by means of a segmented copper cylinder, or disk, called a **commutator.** Commutation is discussed further in the section on d-c generators.

In a-c generators, commutation is not employed at all. The alternating current generated is transferred directly to the external circuit. When the generated current is produced in a revolving armature, it is transferred by means of **collector rings,** or **slip rings.** Slip rings are simply solid copper or brass cylinders, or disks, mounted on the same shaft as the armature of the machine. The alternating current is collected from these rings by means of brushes which maintain contact with the revolving slip rings through the action of springs.

**The Revolving-armature Alternator.** There are, in general, two major parts to any generator—the armature and the field. The **armature** is that portion of a generator in which the emf is induced. The **field** is that portion of a generator that provides the magnetic field. The generation of an emf in the armature is due to the lines of force of the magnetic field cutting the conductors of the armature. Since this is an application of the basic laws for generator action (Chaps. I and V), an emf is induced only when the conductors of the armature are moving with respect to the lines of force of the magnetic field. In the revolving armature alternator this motion is accomplished by permitting the armature to revolve between the poles of a powerful electromagnet.

Figure 56 illustrates an elementary revolving armature alternator. The armature is represented by a single turn coil so arranged that it can revolve about its axes $AA'$ when a mechanical torque is applied to its shaft. In commercial alternators, this coil is wound about an iron core. The greater permeability of iron permits a concentration of the lines of force through the armature resulting in much greater efficiency. In such machines, the iron core is referred to as the armature proper. The coil is referred to as the armature winding.

The two ends of the coil in Fig. 56 are connected to slip rings that are insulated from each other and from the armature shaft.

A powerful magnetic field is maintained across the area through which the armature revolves by two *field poles* marked N and S in Fig. 56. Each of these poles is wound with a number of turns of wire. The coil about the N pole is wound in the opposite direction from the coil about the S

pole. The two field coils are connected in series and connected to an external source of direct current. This is shown as a battery in the illustration. In commercial alternators, with their relatively high power output, a battery would be a very unsatisfactory source of power for this purpose. In commercial installations, therefore, a small d-c generator is utilized for this purpose. Such a generator is called an **exciter,** since it furnishes the excitation power that creates a magnetic field.

Since the field coils are wound in opposite directions, the same direct current flowing through both coils will create magnetic fields which are opposite in polarity. The left-hand pole is a *north* pole, the right-hand pole, a *south* pole for the condition of Fig. 56. As a result, the path of the magnetic lines of force is unbroken by the gap in which the armature revolves. The lines of force travel through the core of one field pole, about the iron housing upon which the field poles are mounted, through the other field pole through the armature and back to the original field pole. This is shown by the dotted line in Fig. 56.

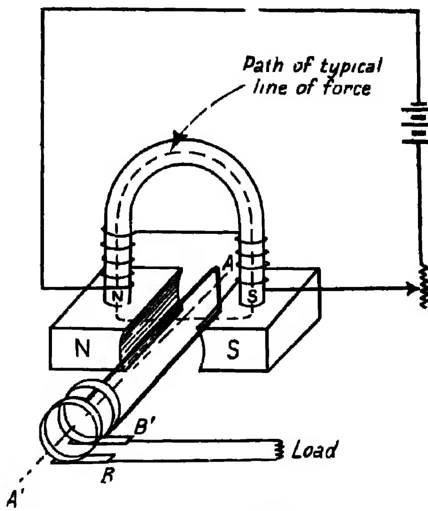



FIG. 56 An elementary revolving armature alternator

So long as the armature is stationary, no emf is induced in its winding, since the magnetic field of the poles is not moving with respect to the armature conductor. When the armature is rotated about its axis, however, the armature winding is continually passing through the lines of force, that is, the magnetic field is *moving* with respect to the conductors of the armature. An emf is therefore induced (an electric field created) in the armature winding. At any instant, the emf present across one horizontal wire of the armature winding is opposite in polarity to the emf induced across the other horizontal wire. The two emf are therefore additive, and the voltage measured across the slip rings (the ends of the single-turn coil) is equal to the sum of the voltages induced in the two halves of the winding.

The generation of a complete cycle of alternating current as the armature passes through 360° of rotation has been discussed in Chap. V and will not be repeated here. The alternating current generated in the winding of the armature is passed to the external-load circuit through the brushes *BB'* in Fig. 56 which maintain contact with the revolving collector rings.

The elementary revolving-armature alternator illustrated in Fig. 56 utilizes only two field poles. In commercial alternators, the output of the machine is greatly increased by using a number of field poles. The field coil windings are so arranged that alternate poles have the same polarity; that is, every other pole is a north pole and the intervening poles are south poles. In such a machine the armature is slotted, dividing the perimeter into a number of poles called **armature poles.** In a given multipolar field machine, there are as many armature poles as there are field poles. The armature coils are wound in the slots in such a manner that all the coils are in series. The armature winding is therefore essentially a single continuous winding. There are as many equally spaced taps on the armature winding connected to the collector rings as there are field poles. Alternate taps are connected to the same ring. Figure 57 shows the armature arrangement for a four pole revolving armature alternator.

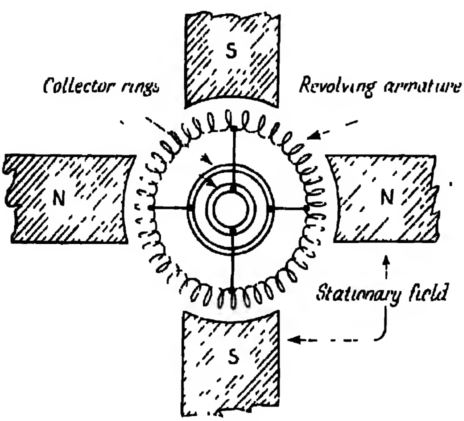



FIG. 57. Four pole revolving armature generator.

The frequency of the alternating current developed by a revolving armature type alternator depends upon the number of field poles and the speed at which the machine is driven. The exact relation is given by the mathematical expression

$$f = \frac{N \cdot S}{120} \qquad (1)$$

where $f$ = frequency in cycles per second,
$N$ = number of field poles;
$S$ = speed of armature in revolutions per minute.

The application of revolving armature alternators is limited in practice to comparatively low power machines. Revolving armature alternators are seldom encountered in capacities greater than 17 kv-amp. The inherent regulation of this type generator is poor for the higher capacities. In addition, it is difficult to insulate armature and collector rings economically for higher voltages.

**The Revolving-field Alternator.** The revolving-field alternator operates upon the same principle of electromagnetic induction as the revolving armature type. In the revolving-field type, however, the d-c excited field is mounted on a revolving shaft. The armature windings are wound

on stationary poles mounted on the generator housing. This type of alternator eliminates many of the disadvantages of the revolving-armature alternator. The only emf that is impressed across the collector rings is the comparatively low value of d-c exciter voltage. The a-c output is taken directly from the stationary armature windings. No moving contacts enter into the output circuit, and the efficiency of this machine

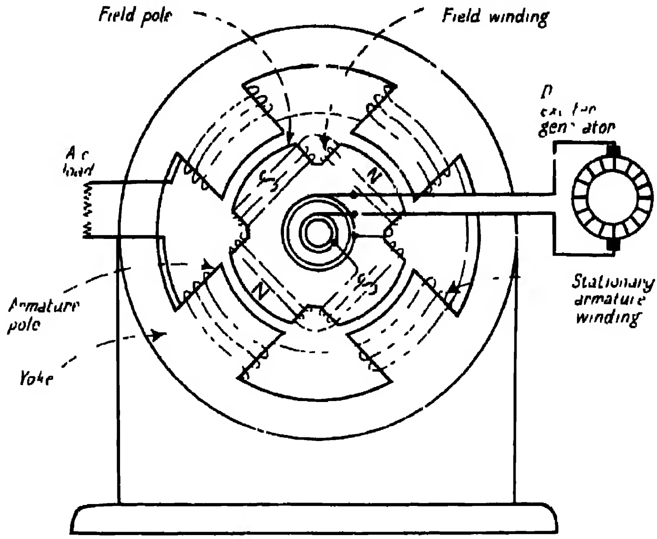



FIG. 58 Elementary four pole revolving field alternator.

is correspondingly greater. A diagram of a four pole revolving-field alternator is shown in Fig. 58.

All modern alternators of large capacity are of the revolving field type.

**The Inductor-type Alternator.** The inductor-type alternator has the tremendous advantage of maintaining both armature and field windings stationary. The moving element is an iron rotor called the **inductor,** which has a number of teeth or poles cut in its perimeter The relative positions of armature field windings, and inductor are such that as the inductor revolves, its poles alternately increase and decrease the intensity of the magnetic field passing through the armature.

An elementary inductor alternator is shown in Fig. 59. It consists of a field magnet externally excited by a source of direct current, a stationary armature mounted opposite the field magnet, and a number of iron disks, or inductors, arranged on a shaft to rotate through the gap between the field and armature poles. When a pair of the iron disks passes through the air gaps between the armature and field poles, the magnetic permeability of this portion of the magnetic circuit is greatly increased. As a

result, the intensity of the magnetic flux through the entire system is greater. As the inductors continue to revolve, the iron disks pass out of the air gap. The permeability of the gap is thereby greatly decreased with resultant lessening of the intensity of the magnetic flux throughout the system.

The variation in density of the magnetic field which cuts the armature windings results in what amounts to a moving magnetic field. Since, according to Lenz's law (Chap. VII), an induced emf is always in such a direction as to oppose the *change* that caused it, the resulting current is

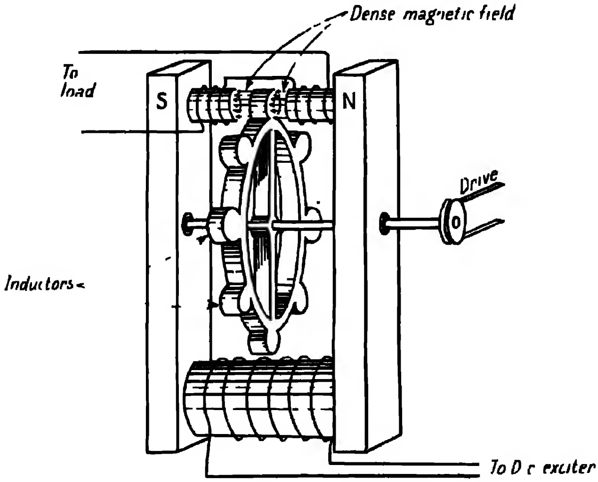



FIG. 59 Elementary inductor alternator.

an alternating current, that is when the magnetic field is *increasing* in intensity, the induced current flows in one direction. When it is *decreasing* in intensity, the induced current flows in the opposite direction.

The essential feature of the inductor alternator is that the magnetic flux is not alternating but undulating in character. With a given magnetic flux through the poles therefore, the total number of armature turns required to produce a given voltage is *twice* that which is required in an alternator having an alternating instead of an undulating flux in its field windings. This large armature is one of the chief disadvantages of the inductor alternator, since it practically prohibits the use of this type of machine in large capacities. Other defects of the inductor alternator are enormous magnetic leakage, heavy eddy-current losses, poor regulation, and ineffective heat dissipation.

For the above reasons, inductor alternators have become practically obsolete except for certain special applications requiring relatively small machines. Nevertheless, inductor alternators at one time found wide application in the field of radio telegraphy. Inductor alternators are

capable of generating alternating currents of higher frequencies than other types of alternators. Machines of this type have been constructed capable of generating frequencies as high as 200,000 c. Many types of spark transmitters utilize inductor alternators that generate 500-c alternating current. Despite the fact that spark transmitters are outmoded, a great many are still in existence, and the student should be familiar with the principles of inductor alternators for this reason.

## *THE D-C GENERATOR*

**Fundamental Theory of Generator Action.** As previously explained, the current induced in the armature windings of a d-c generator, or

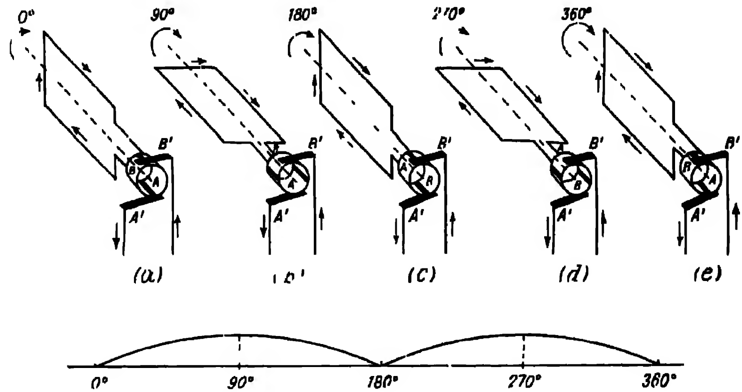



FIG. 60. (a) to (e) The process of commutation in an ordinary dynamo. (f) Output wave form for the commutation cycle.

dynamo, is alternating current, just as in the armature windings of an alternator. The distinction between the two types of machines lies in the manner of taking the current from the armature. In the dynamo, the alternating current of the armature is converted into direct current by the process of *commutation*.

Commutation in an elementary single-turn armature is illustrated in Fig. 60. The process consists of placing a form of revolving switch, called the **commutator,** between the armature and the external circuit. This switch is so arranged that it will reverse the connections with the external circuit at the instant of each reversal of current in the armature. Commercial commutators are composed of a number of copper bars, or segments, arranged side by side to form a cylinder. The segments are insulated from each other and from the armature shaft upon which they are mounted by sheets of mica or other high grade insulating material. The commutator of the elementary dynamo illustrated in Fig. 60 has but two segments.

As the single-turn armature winding of Fig. 60 rotates in a clockwise direction, the current delivered to the external circuit is shown graphically in Fig. 60(f). As the armature rotates from its starting position (a) through a quarter revolution (90°) to position (b), the current rises from zero to maximum amplitude and flows in the direction indicated by the arrows. The current passes to the external circuit through the commutator and its brushes, leaving through segment *A* and brush *A′*, and returning through brush *B′* and segment *B*. As the armature continues its rotation through an additional 90°, the current, although still flowing in the same direction, falls from maximum amplitude to zero at position (c). The current is passed to the external circuit through the same commutator segments as during the first 90° of rotation.

At the beginning of the second half of the revolution, the current in the armature reverses its direction as indicated by the arrows in position (c). The rotation of the commutator simultaneously reverses the contacts between commutator segments and brushes. Segment *A* is now making connection with brush *B′*, and segment *B* is making contact with brush *A′*. The current in the external circuit, therefore, still flows in the same direction as during the first half of the revolution. As the armature rotates through the remaining half revolution, the current rises to maximum amplitude again at the 270° position (d) and falls to zero at the 360° position shown at (e). The armature is now back at the original starting position, since position (a) and (e) coincide. At this point, the contact between commutator segments and brushes is again reversed, and another cycle of generation commences coincident with the next revolution.

It is at once apparent that the output of a single-turn dynamo such as that of Fig. 60 is pulsating. The current not only rises to maximum values, but also falls to zero twice during each revolution. The ideal direct current is a continuous current, that is, one that does not fluctuate in amplitude. Actually, a continuous current is never obtained from a dynamo. Nevertheless, the ideal is approached by winding the armature with a great number of coils instead of with a single turn, as in Fig. 60. The coils are so connected to the commutator segments that the successive coils begin the cycle progressively. Figure 61 is a graph of the output current of a multiwinding dynamo. As the number of coils in the armature is increased, the amplitude of the pulsations decreases so that the resultant curve approaches the form of a straight line. In commercial dynamos, there are a great many coils, and it is possible to confine the amplitude of the pulsations to an exceedingly small value, which is not objectionable for most applications. Such a current for all practical purposes may be referred to as a continuous current. The amount of pulsation that *is* present is generally referred to as **commutator ripple.** In some radio applications, it is necessary to remove even this small

ripple. The removal is accomplished by means of inductance-capacitance filters of the type described in Chap. XII.

There are many different types of dynamo armatures, including the *disk*, *ring*, and *drum* types. The drum-type armature is used almost to the exclusion of the other types in this country. The drum-type armature is comprised basically of a cylindrical core usually drilled with a number of longitudinal holes for ventilating and cooling purposes. The windings are disposed on the cylindrical surface of the core, usually in open-type longitudinal slots. In contrast to the ring-type armature, in which the coils are wound in and out of the ring-shaped core in helical fashion, no

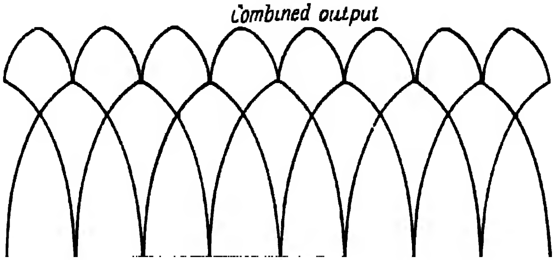



FIG. 61 Output wave form of multiwinding dynamo

part of the winding in the drum armature is permitted to thread through the core.

All armature windings may be classified in two general types, namely, the open coil type and the closed-coil type. **Open-coil** windings are used most generally on a-c machines. Such a winding can be identified by starting with any conductor and tracing progressively through the winding until a "dead end" is finally reached. If a conductor of a **closed-coil** winding is traced, the starting point will always be reached eventually. The entire winding, or some submultiple of it will be traversed before reaching the starting point again. Armature winding classifications may be further subdivided into such types as single and double layer, lap, wave, spiral, and fractional slot.

**The Series-wound Dynamo.** Dynamos of the type used in most radio work are **self-excited dynamos,** so called because a portion of the direct current generated by the dynamo itself is fed back to the field windings for excitation. Such generators require no external source of excitation. The initial excitation of the field is due to the residual magnetism retained by the iron cores of the field poles. Often, a dynamo loses its residual magnetism and will fail to build up an output voltage. In such cases, it is necessary to apply a d-c voltage from an external source to the field windings to remagnetize the field cores. The external voltage can be a very small one and need be applied only momentarily to accomplish the remagnetization.

When a self-excited dynamo is first turned over, the emf generated in the armature by the field on account of its residual magnetism is very small. The small output current flowing through the field windings increases the magnetic field. This in turn increases the output. The action is progressively cumulative, and the generator builds up to full output voltage in a few seconds.

There are three methods of connecting the field windings in dynamos. In the series-wound dynamo, the field windings are connected in *series* with the output circuit (the armature) The connections of an elementary bipolar series-wound dynamo are shown in Fig. 62, where there is one

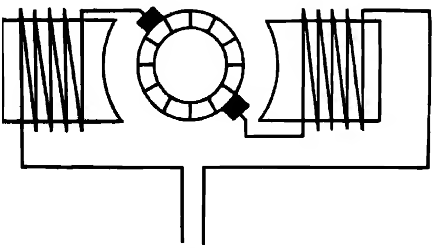



FIG. 62 Series-wound dynamo

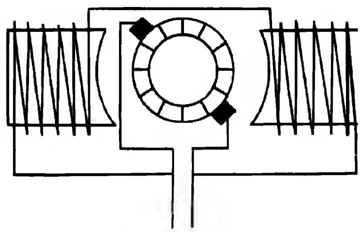



FIG. 63. Shunt wound dynamo

field winding in each leg of the output circuit. When the circuit is closed, that is, connected to a load, the current developed by the armature passes through each field coil as well as the load. Since *all* the current generated by the armature must pass through the field windings, it is necessary that these windings must be wound of large wire in order to carry the current safely. The necessary field strength is therefore developed with comparatively few turns of wire

A series-wound dynamo will furnish current at increased voltages as the load increases up to a certain point. If sufficient current is drawn to overload the machine, the voltage will drop.

**The Shunt-wound Dynamo.** In this type of dynamo, the field windings are shunted, or paralleled, across the output of the generator. Figure 63 illustrates the connections of a shunt-wound dynamo. The excitation circuit, or field windings, comprise an independent circuit and can be thought of as constituting an additional load upon the armature connected in parallel with the regular load. The resistance of the field circuit must therefore be kept very high in order not to short-circuit the load circuit The necessary field strength is built up by utilizing many turns of relatively small wire. Consequently, only a small portion of the current induced in the armature is absorbed by the field windings Since the shunt-wound dynamo is also a self-excited generator, it depends upon residual magnetism for its initial field.

The shunt-wound dynamo is characterized by poor voltage regulation. There is a certain maximum-load current that a given shunt-wound dynamo is capable of supplying at constant voltage, but beyond this value, the voltage of the generator rapidly decreases as the load increases.

**The Compound-wound Dynamo.** The compound-wound dynamo amounts to a combination of the series- and shunt-wound machines. It possesses some of the characteristics of each type machine and is designed to provide automatically better regulation than is possible with either the series or shunt-type generators.

The compound-wound dynamo is wound with two sets of field coils,

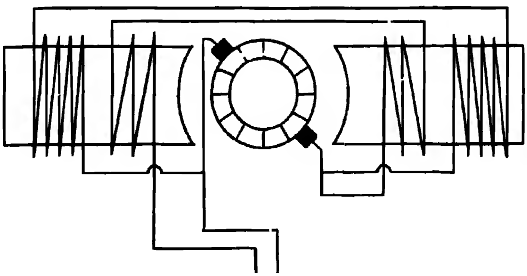



Fig. 64. Compound wound dynamo

one set connected in series and the other in parallel with the armature. A diagram of a compound wound dynamo is shown in Fig. 64. The number of turns and relative resistances of the series and shunt windings are so adjusted that the voltage at the output terminals is maintained practically constant over a wide range of loads.

Dynamos are utilized in many types of radio stations as the source of d-c plate voltage for transmitters. Often, in such applications, h-f (r f) voltages from the transmitter generated by the vacuum tubes are induced in the generator output leads and find their way back to the dynamo itself. Such voltages, if high enough, could puncture the insulation of armature or field windings in the machine, but since they are alternating voltages, they are usually by-passed to ground by capacitors which are installed as protective devices across the brushes of the dynamo. The capacitors afford a low resistance path to ground for the undesired h-f voltage. The d c output of the dynamo is unaffected by their presence.

Very often the sparking at the brushes of a dynamo used in a radio station causes undesirable interference in radio receivers. If the receiver is powered by a dynamo, the ripple component (commutator ripple) of the generator output may cause interference. Interference from either sparking at the brushes or ripple component may usually be eliminated by the insertion of an inductance-capacitance filter having the proper circuit constants. Two such filters, each with different component

values, are usually necessary to suppress interference from two different sources. The design of such filters is discussed at length in Chap. XII.

The voltage of a generator, either alternating or direct current, is regulated by varying the amount of excitation supplied to the machine. Since alternators are always externally excited, variation is accomplished in these machines by inserting a rheostat (variable resistor) in series with the exciter generator output and the field windings. In dynamo circuits, the type of excitation control varies with the type of dynamo.

The series-wound dynamo voltage output is regulated by a rheostat connected in parallel across the series field winding. As the resistance of the rheostat is increased, the amount of current flowing through it is decreased. The current flowing through the field therefore increases, increasing the voltage output of the machine. As the rheostat resistance is decreased, the current flowing through the field winding is decreased, resulting in lower voltage output.

In the shunt-wound dynamo, the power rheostat is inserted in series with the armature and the field windings, and the action is the same as for externally excited machines. In the compound wound dynamo the rheostat is also connected in series with the armature and the shunt winding. A rheostat used in this manner to control the output voltage of a generator is called a **field rheostat.**

## *THE A-C MOTOR*

The electric motor finds a multitude of applications in the modern radio station. In addition to providing the motive power to turn over generators of various types, it is used to drive pumps in transmitting tube water-cooling systems, for the operation of air cooled tube blowers, and for any number of other important applications. Since the electrical distribution systems aboard ships are d c systems, the d c motor finds its greatest application in mobile radio stations. At coastal and broadcasting stations, the source of power is alternating current, and a c motors, therefore, are used practically exclusively at fixed radio stations.

Alternating-current motors may be divided into two general classifications synchronous and asynchronous (nonsynchronous) motors. The term "synchronous" means "in unison," or "in step," and a **synchronous motor** may therefore be defined as one that rotates in unison or in step with the phase of the alternating current operating it.

Synchronous motors are especially suited for high-voltage service and are desirable in applications requiring large amounts of power. The speed of the synchronous motor depends upon the frequency of the alternating current and the number of poles and is independent of other factors. If the load becomes too great, the motor falls out of phase and will stop. Such motors require an auxiliary power for starting. For

low-power applications and for applications where the static load is heavy, single phase synchronous motors are usually undesirable. For these reasons, such motors find very little application in the radio field and will not be discussed further here.

**Asynchronous Motors** can be divided into induction motors and commutator motors. The commutator motors can be further divided into three types: series, compensated, and repulsion type commutator motors.

The superiority of the repulsion-type motor to other types soon led to its almost exclusive use for small-motor applications. For this reason, only this type of a-c motor will be discussed in this chapter. The repulsion-type commutator motor is often referred to as the **repulsion-induction**

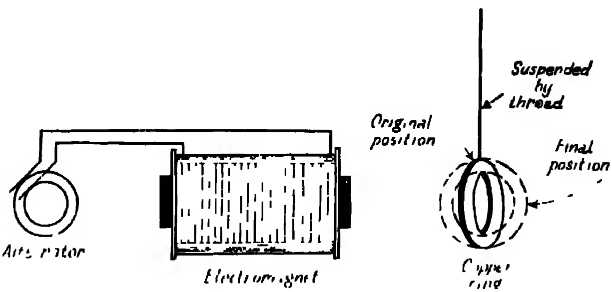



FIG. 65. Principle of the repulsion induction motor

motor a name derived from the fact that the repulsion principle is utilized only for starting the motor. Once started, the motor then operates as a simple induction motor.

**The Repulsion-induction Motor.** The principle underlying the operation of this type was discovered first by Elihu Thomson in 1887 during the course of his experiments with alternating currents. Thomson found that if a closed coil, such as the copper ring in Fig. 65, is suspended in an a-c field, an emf will be induced in the coil that will be 90° out of phase with the inducing flux. Since the ring is a closed circuit, current flows in the ring, and as a result, another magnetic field is created by the current in the ring. These two fields are always in such a direction as to repel each other. If the copper ring is so suspended that its plane is oblique to the lines of force, the repulsion of the two fields will cause it to turn until its plane is parallel to the lines of force. This principle of repulsion is utilized to supply the original torque of starting to a simple induction motor.

The induction motor depends for its operation upon a reversing magnetic field. The induction principle can best be understood by referring to Fig. 66. The two field poles are so wound that they have opposite magnetic polarity at any instant. An armature having a single-turn coil

is so mounted between the field poles that it is free to rotate. If the armature coil is in an *oblique* position between the fields, the magnetic field produced by the current induced in the coil will cause the coil to rotate. If the coil is *not* in an oblique position, the forces exerted by the two fields will be equal and opposite, and no current will be induced in the coil. The alternating current causes continuous reversals of the field polarities, with the result that magnetic forces are always exerted upon the coil and cause it to continue to rotate.

A simple repulsion motor consists of an armature, a commutator, and field poles. The armature is wound in exactly the same manner as is the armature of a d-c generator. The brushes are placed about 60 or 70° from the neutral axis and all are connected by heavy short circuiting bars. With this arrangement, only the coils of the armature that are connected with the brushes at any instant form a complete circuit. Since the brushes are not in the neutral axis, these coils are sure to be in an oblique position with respect to the field poles. The motor can therefore *always* be started, regardless of the position of the armature. Once started, such a motor, if nothing is done to prevent it, will increase in speed at no load until the armature bursts. This result is prevented, however, by a governor arrangement. When the motor has attained speed, the governor operates to short circuit all the armature windings usually by means of a necklace arrangement on the commutator facing. The motor then runs as a squirrel-cage induction motor, that is, all the armature coils are in the circuit and have emf induced in them. Most modern motors have an arrangement that lifts the brushes off the commutator when the armature is short circuited, thus prolonging the life of both the commutator and the brushes.

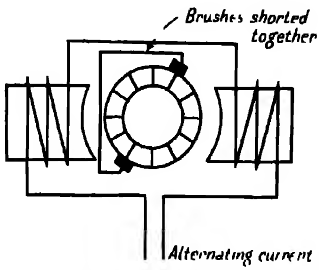


FIG. 66. Elementary two pole induction motor.

## THE D-C MOTOR

There are three basic types of d-c motors: the shunt wound motor, the series-wound motor, and the compound-wound motor. A special kind of the latter type is the differential compound-wound motor. There is no essential difference in the construction of d-c motors and generators; the function of the machine is simply reversed.

Certain principles of operation are very important for the explanation of the operation of d-c motors. The useful force on a conductor in a magnetic field is proportional to the product of the current in the conductor, the magnetic-field strength, and the cosine of the angle between

the direction of motion of the conductor and the direction of the magnetic field. The total torque delivered by a motor is proportional to the product of the force on each conductor, the number of conductors, and the distance from the conductors to the center of rotation of the armature. As the conductor rotates in the magnetic field, a back emf is produced, which tends to oppose the current in the conductor. This back emf is proportional to the velocity of rotation and to the magnetic-field strength, and the difference between the applied emf and the back emf determines the current in the conductor. In other words, the current in an armature winding is equal to the emf from the external source of power minus the back emf divided by the resistance of the winding.

**The Shunt-wound Motor.** A shunt wound motor is one in which the field coils and the armature are connected in parallel, or shunt, and each receives the same applied emf from the external source of power. If the applied emf is constant, the magnetic-field strength through the armature is a constant. The force on the armature windings is always in the same direction because of the commutating action, as described for the case of the dynamo.

Starting from rest, the armature will continue to pick up speed until the back emf has reduced the armature current to the point where the resulting torque is equal to the torque required by the load. It can be seen that the shunt wound motor will not exceed a certain speed, even with no load and no friction in the motor. This top speed is the speed at which the back emf is equal to the applied emf and no current flows in the armature

The shunt-wound motor is characterized by reasonably constant speed for all loads within the capacity of the motor.

**The Series-wound Motor.** A series wound motor is one in which the field coils, consisting of relatively few turns of heavy wire, are connected in series with the armature of the motor. All the current supplied to this machine therefore passes through the field coils as well as through the armature.

Series wound motors have comparatively large starting torques, because at low speeds, when the back emf is small, the large armature current permits a large current to flow and, hence, creates a strong magnetic field

Series-wound motors are characterized by relatively poor regulation, the motor speed changing considerably as the load is varied. If a series-wound motor with small bearing friction is run at no load, its speed will continue to increase until its armature bursts from centrifugal force. Under the no-load condition, the speed tends to increase to the point where the back emf is equal to the applied emf. Since the field is in series with the armature, however, the reduced armature current also means a reduced magnetic field, so that the armature must revolve still

faster to produce sufficient back emf, and the process can continue indefinitely.

**The Compound-wound Motor.** The construction of this motor is very similar to that of the compound-wound generator. This motor combines some of the characteristics of both the shunt-wound and the series-wound motors. The series winding gives the motor a strong torque at starting, and the shunt winding prevents excessive speed.

The regulation of the compound-wound motor, however, is not quite so good as that of the shunt wound motor.

**The Differential Compound-wound Motor.** The field and armature connections of this type of motor are the same as those of the simple compound-wound motor, except that the series and shunt fields are wound so that their fields oppose, or buck, each other. The magnetizing force of the shunt winding is normally the greater, but as the load on the motor increases, the increased series-field current tends to reduce the total field strength, which reduces the back emf. The increase in armature current necessary to supply the additional torque is therefore brought about by a reduced magnetic field rather than by a reduction in speed. The differential compound wound motor may be arranged to operate at an almost constant speed for a certain range of load conditions.

## *THE MOTOR GENERATOR*

The term **motor generator** is applied to any combination of an independent motor and an independent generator whose shafts are coupled. Such units are usually mounted on a common steel base in order to provide secure alignment of the machine bearings.

Motor generators find a variety of applications in radio stations. They are used aboard ships as a convenient means of converting the direct current of the ship to alternating current to supply a radio transmitter. Such a unit, of course, would consist of a d-c motor directly coupled with an alternator. The field of the alternator is usually excited directly from the d-c mains of the ship

Motor generators are often used at broadcast or coastal radio stations to supply direct current from the alternating current of the city power system. Such units consist of a d-c generator driven by an a-c motor and are used when it is impractical to rectify the alternating current by means of vacuum-tube rectifiers as described in Chap. XII. As a rule, such motor generators are used when the current requirements are very heavy and vacuum-tube circuits would be relatively inefficient.

## *THE DYNAMOTOR*

A dynamotor differs from a motor generator in that the motor armature and the generator armature are combined into one, thereby requiring

but one field frame. Since the motor and generator armature windings are mounted on a single core, the armature reaction due to one winding is neutralized by the reaction caused by the other winding. Consequently, there is little or no tendency for sparking to occur at the brushes.

Dynamotors are usually bipolar machines having only shunt windings on the field poles. There are two sets of windings on the armature connected with commutators at opposite ends of the shaft. In general, the function of the dynamotor in d-c circuits is similar to that of the transformer in a-c circuits. Dynamotors are used to convert d-c power at one voltage and current to d-c power at a different voltage and current. The ratio of output to input voltage, or current, depends upon the inherent construction of the dynamotor. For a given machine, these values are not variable functions.

Originally, the dynamotor was designed as a combination of a d-c motor and a dynamo, or d-c generator, and its primary application was in d-c circuits. The name **dynamotor** was thus derived from the application. In more recent times, however, this machine was utilized in both a-c and d-c circuits to convert electric power at one voltage and current to power at another voltage and current, regardless of whether the input and output were alternating or direct current. The name dynamotor is still applied to such machines. Thus, a dynamotor can be used to convert *alternating current* at a given voltage to either *direct* or *alternating current* at a higher or lower value of voltage. Similarly this machine can be used to convert *direct current* at a given voltage to either *alternating* or *direct current* at a higher or lower value of voltage.

The major difference in dynamotors for the various applications outlined above lies in the method of feeding and collecting current from the armature windings. Thus, a machine used to convert alternating current to alternating current of a different voltage value will have slip rings on the motor side and also on the generator side. Actually, dynamotors are very seldom employed for this particular application, since this function is much more efficiently performed by the transformer. If a dynamotor is used to convert direct current to direct current at different voltage and current values, the machine will be equipped with commutators at both motor and generator ends. If used to convert direct current to alternating current it will have a commutator at the motor end and slip rings at the generator end.

The dynamotor can be used either to step up or to step down circuit voltages, regardless of whether they are alternating or direct voltages. This ability depends upon the ratio of the turns of the generator and motor armature windings. Thus, a step-up dynamotor will have a greater number of turns in the generator armature winding than in the motor armature winding. If the machine is a step-down unit, the opposite condition will prevail.

In general, the dynamotor is distinguished from other types of combination machines in that a common armature core having *two* armature windings is used.

## THE ROTARY CONVERTER

The rotary converter is an important modification of the dynamotor and is used wherever it is desired to convert alternating current to direct current, or vice versa. The machine has only *one* armature winding, which is common to both the motor and generator ends of the machine. Because of this constructional characteristic, the rotary converter cannot be used as a means of stepping-up or stepping-down circuit voltages.

A rotary converter is inherently a reversible machine; that is, if it is supplied with direct current at the proper voltage at its commutator end, it will run as a d c motor and will deliver alternating current to the collector rings. If it is supplied with alternating current at the proper voltage at its collector ring end, it will run as a synchronous a c motor and will deliver direct current from its commutator end.

The main advantage of the rotary converter is its low initial cost as compared with either a motor generator or dynamotor performing the same function. The dynamotor has two armature windings and is therefore more expensive to construct than the rotary converter which has only one. The dynamotor, however, is a more efficient machine than the rotary converter for the same application. The motor generator, although more expensive to construct than either the dynamotor or the rotary converter, is more efficient than either type.

## BRUSHES

Brushes are utilized in almost all types of motors generators, and combinations of motors and generators and in any type of machine where it is desired to make electrical contact with a revolving surface Good brushes are required to have current-carrying capacity sufficient for the particular application; they must be good conductors to ensure low-voltage drop and must be constructed of a material that will adapt itself to wear, so that close contact will be maintained with the revolving surface.

One of the most commonly used materials for brushes in motors and generators is carbon. Carbon is a comparatively good conductor and, being soft, wears easily. This wearing characteristic of carbon causes brushes made of this material to wear down soon to the exact shape of the commutator with which it is making contact. The good contact that results reduces sparking at the brushes and insures minimum electrical losses at this point. As the carbon continues to wear, it disintegrates in the form of a very fine powder and does not short-circuit adjacent

commutator segments if it lodges in the mica insulation between segments. Carbon is used for brushes in practically all applications where the current requirements do not exceed 35 amp. Carbon is not capable of carrying currents in excess of 35 amp per square inch.

In multipolar machines, where the current exceeds 35 amp, brushes constructed of copper leaf are utilized. A number of sheets of copper leaf, or foil, are compressed to form a single brush and the edges of the leaves are brought to bear against the commutator or slip ring surface. Owing to the thinness of the individual leaves, such brushes have fairly good wearing qualities and soon wear down to the shape of the commutator or slip ring. In addition, these brushes have the advantage of a current-carrying capacity in the vicinity of 140 amp per sq. in. Commutator-type machines utilizing copper-leaf brushes should be subjected to frequent inspection, since minute particles of copper caused by brush wear may find their way into the crevices between adjacent segments. Sparking, excessive brush and commutator wear, and other troubles may result from this condition.

## *CONTROL CIRCUITS*

There are two general types of control circuits used in conjunction with motors and generators, namely, starting circuits and protective circuits. Actually, every commercial installation combines some type of each of these circuits. All these control circuits are designed to protect the machine in question from the harmful effects of improper starting, overload, or underload

Starting circuits may be divided into two classes, namely, ***hand-starting*** circuits and *automatic starting* circuits. Some sort of starting device is absolutely necessary in circuits where the motor has appreciable capacity. A motor armature has a low static resistance. If an attempt is made to start a motor by directly connecting it across the voltage source, excessively high current will flow because of the very low armature resistance and the absence of back emf. Such quick starting could result in burning out the armature winding. Once the motor has built up speed, however, the back emf built up in the armature safely limits the current flow, and the motor can be directly connected across the power mains.

Safe starting of motors, including devices utilizing the motor principle (dynamotors, rotary converters, and so on), is accomplished by inserting a resistance in the line. The value of resistance, of course, varies with the type of motor with which it is used. In general, the resistance should be of a value that will limit the flow of current in the armature windings to a safe value. When the circuit is closed with this resistance in the line, the motor will revolve very slowly because of the limited amount of current flowing through the armature. Once the motor is revolving, however, a

back emf begins to build up in the armature winding and tends to limit the current. The armature is therefore no longer so susceptible to harm from higher currents. Accordingly, some of the resistance is cut out. The motor then increases its speed, and the back emf increases. The resistance is taken out of the circuit in progressive steps until finally the motor is operating directly across the power mains and is running at full speed.

**Hand-starting Apparatus.** An elementary circuit of a hand-starting motor control is shown in Fig. 67. The movable lever $L$ is equipped with a handle to enable the operator to manipulate it. At the far left-hand position, the lever is not making contact with any of the resistor contact positions, and no current reaches the armature. As the lever is moved in a clockwise direction, the first resistor is contacted, thus closing the

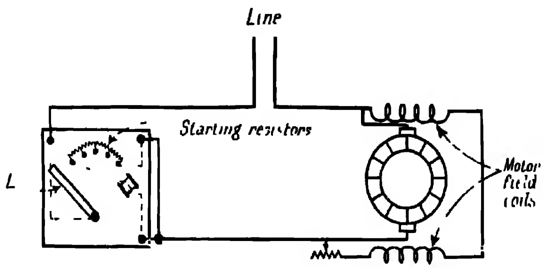



FIG. 67. A motor hand starter circuit.

circuit between armature and power mains. The full resistance of the starter is in series with this circuit. As the movement of the lever is continued in a clockwise direction, successive sections of the starting resistor are cut out of the circuit, until finally, with the motor running at full speed, the armature is connected directly across the mains.

Hand starters are always equipped with *no-voltage release coils*. These coils are electromagnets connected in series with the motor field and the source of excitation. Aboard ships, where the same d c source is applied to the armature and the field, the return circuit of the no-voltage release coil is to one side of the power mains. This electromagnet acts upon the iron starter handle and fulfills a double purpose. Once the motor has attained full speed, the field current flowing through the electromagnet causes it to hold the starter handle on the full "on" position. If, for any reason, the power source is disconnected from the motor, the electromagnet is de-energized, and the starter handle is quickly pulled back to the open-circuit position. This operation prevents the accidental application of full current to the armature after the circuit has been interrupted.

Care should be exercised in operating a motor hand starter. If the starter is moved too quickly, excessive current will flow through the armature windings and may burn them out. If the starter is operated too slowly, current may flow through the starting resistors for too long a

time, and they may burn out. Starting resistors are designed for temporary duty and will not carry heavy current for any appreciable length of time.

**Automatic-starting Apparatus.** There is a wide variety of automatic-starting devices. In general, automatic starters perform the same function as hand starters. The difference lies in the fact that the resistors are cut out of the circuit automatically, and the time element between the operation of successive contactors is therefore fixed. For this reason, automatic starters are relatively more efficient than hand starters.

In general, automatic starters consist of a multiposition contact bar similar to the lever in a hand starter and operated by a solenoid. The operator actuates the mechanism by pushing a button that closes a circuit with the solenoid winding. The movable iron core in the solenoid is set in motion by the attraction of the magnetic field created by the solenoid. A contact bar is fastened to the iron plunger in such a manner that, as the plunger moves, a series of contacts connected to the starting resistors are successively engaged by the contact bar. In some types of automatic starters, the iron plunger moves through a container filled with heavy viscous oil that retards the speed with which the plunger moves. Such starters can be adjusted so that the time element may be varied from 5 to 20 sec.

**Protective Devices.** Protective devices may be classified as overload and underload devices. The most common type of overload safety device is the fuse, which is so well known and widely used as to require no explanation here. All the overload and underload devices used in connection with starting circuits are of the type that operate to open the circuit. For this reason, they are customarily called **circuit breakers** and are classified according to function. Thus an overload circuit breaker is one that will automatically open the circuit if the current value becomes too high, and the underload circuit breaker is one that will automatically open the circuit if the current value becomes too low.

An overload circuit breaker consists, basically, of a manually operated switch, the contacts of which are in series with the motor control circuit. When the switch is closed, it is kept in this position by means of an iron-bar latch. The latch is mounted on a pivot so that its opposite end is in the field of an electromagnet, which is in series with the control circuit. When the switch is closed, therefore, current flows through the electromagnet as well as through the motor-starting circuit. The spacing between the iron-bar latch and the electromagnet is adjustable and for normal operation is fixed at a point where the normal current flowing through the electromagnet is not sufficient to trip the latch. If, for any reason, excessive current flows through the circuit, the field of the electromagnet increases. As a result, the iron-bar latch is attracted to it, thereby

releasing its hold on the switch lever, and a powerful spring immediately throws the switch open, breaking the circuit.

The main advantage of an overload circuit breaker is that, unlike a fuse, it may be used over again indefinitely. All that is necessary is to reclose the switch by hand, thus restoring the original position of the mechanism. A circuit breaker, once having tripped, should never be reset until the cause of its tripping has been ascertained and corrected.

In an underload circuit breaker, an electromagnet acts to hold a manually operated switch closed when contact is made. The iron-bar switch armature is attracted by the field of the electromagnet against the action of a spring. The spring tension is adjusted to a point where any decrease in current through the electromagnet will decrease its field enough for the spring to overcome its attraction and open the circuit.

## QUESTIONS AND PROBLEMS*

**1.** What is the purpose of a commutator on a d-c motor? On a d-c generator?

**2.** If a self-excited d-c generator failed to build up to normal output voltage when running at normal speed, what might be the cause, and how could it be remedied?

**3.** Why is carbon commonly used as a brush material?

**4.** How may the output voltage of a separately excited a-c generator at constant output frequency be varied?

**5.** Why are by-pass condensers often connected across the brushes of a high-voltage d-c generator?

**6.** What may be the trouble if a motor generator fails to start when the starter button is depressed?

**7.** Describe the action and list the main characteristics of a shunt-wound d-c motor.

**8.** What determines the speed of a synchronous motor?

**9.** What is the output frequency of a generator having 10 poles and revolving at 1,200 rpm?

**10.** How may the r-f interference that is often caused by sparking at the brushes of a high-voltage generator be minimized?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

# Chapter VII

# INDUCTANCE

It was seen in an earlier chapter that an electric current, or a moving electric field, produces a magnetic field. Conversely, a magnetic field that is moving or changing in intensity produces an electric field, or induced emf.

When current flows in a conductor, a certain amount of energy is stored in the magnetic field surrounding the conductor. When the flow of current ceases, the energy flows back into the conductor, and the magnetic field is said to collapse.

## *SELF-INDUCTANCE*

When the rate of current flow in a conductor is changing, the magnetic field surrounding the conductor is changing in intensity, and, as a result, an emf is induced in the conductor. Lenz's law states:

*The electromotive force produced by electromagnetic induction is always in such a direction that it opposes the change in the original electric field which produced it.*

That is, the induced emf acts in a direction to oppose the change of current. The emf so produced is called the **back emf of induction,** and the ability of a circuit to produce such an emf is called the **self-inductance** of the circuit.

The self inductance, or, simply, inductance, of a given conductor depends on its configuration. A piece of wire will have a greater inductance when wound in the form of a coil than when it is straight because a part of the magnetic field of each turn of the coil surrounds the other turns. The inductance of a given conductor may also be varied by varying the permeability of the medium surrounding it. Thus, if an iron core is inserted in a coil, its self-inductance is increased many times. Once the surrounding medium and the shape and form of a conductor are fixed, the inductance is a constant.

The unit of inductance is the **henry,** so called in honor of the American scientist Joseph Henry because of his many important discoveries in the field of magnetism. *A circuit is said to have a self-inductance of one henry when a current changing at the rate of one ampere per second induces an electromotive force of one volt in the circuit.* The induced emf in a circuit is therefore equal to the inductance of the circuit in henrys multiplied by the rate of change of the current in amperes per second.

Conversely, the rate of change of current in amperes per second in a circuit is equal to the emf in volts impressed on the circuit divided by the inductance in henrys.

Assume that an alternating voltage is impressed on a circuit that contains self-inductance and no resistance. Figure 68 is a graph showing the sine wave of impressed emf and the resulting current flow for such a circuit. When the emf is at a maximum, the rate of change of current is at a maximum. This relation is the point of greatest slope on the current sine wave, which is the point at which the current is zero. When the emf is zero, the rate of change of current is zero, which is the point of maximum current. Thus, it can be seen that the current *lags* the voltage by 90°, or the voltage *leads* the current by 90°.

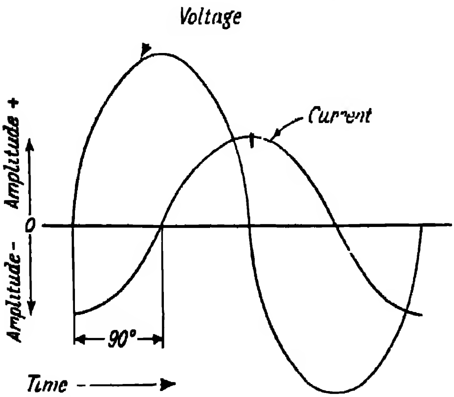



FIG. 68. Voltage and current relations in a pure inductive circuit

Actually, every circuit is made up of both resistance and self inductance, so that Fig. 68 represents a theoretical condition. When resistance is present, the voltage across the circuit may be divided into two parts: a voltage that is equal to the product of the resistance and the *instantaneous* value of current and a voltage that is equal to the product of the inductance and the *rate of change* of current. These two components of voltage and their sum are shown in Fig. 69. Note that the total emf lags the current by some angle that is between 0 and 90°.


FIG. 69. Resistive and inductive voltage components in an inductive circuit containing resistance.

Since the back emf of self-induction is generated only when the current through a conductor is *varying* in amplitude, it is apparent that the flow of direct current through an inductive circuit will produce no inductive effect, that is, no back emf. Since a direct current does *not* vary in amplitude,

the magnetic field produced by it will not be a *moving* field. Nevertheless, there are certain inductive effects connected with direct currents. When an inductive d-c circuit is first closed, the current rises from zero to the maximum value of the current determined by the resistance of the circuit. During the time in which the current is rising, it is, of course, varying in amplitude, and a back emf of self-induction is generated. Similarly, when the circuit is opened, the current in falling from maximum value to zero is again varying in amplitude. Consequently, a back emf of self-induction is again generated. The resultant lag of the current behind the voltage accounts for the greater spark across switch contacts when such a circuit is opened. The lag also accounts for the delay in reaching maximum amplitude when the switch is first closed in such a circuit. This time delay may be very great in circuits possessing much inductance. Thus, in a field circuit of a very large generator, for example, the current may take several seconds to reach its final value. After the current has reached its final ($E/R$) value, it remains at this constant amplitude for as long as $E$ and $R$ remain constant. During this period, the current is constant, and there are no inductive effects whatever, that is, no back emf is generated.

## *INDUCTIVE REACTANCE*

**Reactive and Resistive Opposition.** Since the back emf of self-induction tends to oppose any change in the amplitude of the current that produces it, apparently any current that varies in amplitude encounters considerable *opposition* in an inductive circuit. The peculiar type of opposition offered by inductance is called **inductive reactance.** In order to handle inductive reactance quantitatively, the reactance is expressed in *ohms* in common with all other types of opposition to electric-current flow. The ohm of reactance should not be confused with the ohm of resistance. A resistive ohm is opposition to current flow of the simplest kind and does not affect phase relations in the circuit. The ohm of inductive reactance, however, in addition to offering opposition to the flow of current, has the added effect of displacing the original phase relations of the circuit, that is, it causes the current to lag the voltage.

As the frequency of an alternating current flowing through a conductor is increased, the rate of change of current is also increased. Obviously, the counter emf produced by this current is increased with the result that the inductive reactance of a given conductor increases with an increase of frequency. It is plain, therefore, that inductive reactance is a function not only of the inductance of the circuit but also of the frequency of the current. This relationship is expressed specifically by the mathematical formula

$$X_L = 2\pi f L, \tag{1}$$

where $X_L$ = inductive reactance in ohms;
$\pi$ = the constant 3.14;
$f$ = frequency in cycles;
$L$ = inductance in henrys.

For most work, an accuracy of two decimal places (6.28) is sufficient for the constant $2\pi$.

**Problem.** If a coil has an inductance of 100 mh, what will its reactance be at a frequency of 1 kc?

**Solution.** The given values of the problem are prepared for substitution in Eq. (1) by reducing them to the units required

$$L = 100 \text{ mh} = 0.1 \text{ h},$$
$$f = 1 \text{ kc} = 1{,}000 \text{ c}$$

Substituting in Eq. (1),

$$X_L = 6.28 \cdot 1{,}000 \cdot 0.1, \tag{2}$$
$$X_L = 628 \text{ ohms}. \tag{3}$$

Inasmuch as inductive reactance varies with the frequency, it is apparent that a given coil will have different reactances at different frequencies. Since, according to Eq. (1), inductive reactance is *directly proportional* to the frequency, it becomes greater as the frequency becomes greater. At radio frequencies, this characteristic becomes very important. A coil that performs efficiently at low frequencies may have such a tremendous reactance value at radio frequencies as to become entirely useless for a given purpose. At the ultrahigh radio frequencies, even a lineal conductor has sufficient inductive reactance so that it cannot be neglected. This factor requires the utmost consideration in the layout of apparatus at high frequencies.

**The Choke Coil.** Since inductive reactance is present only when a current of *varying* amplitude is flowing through a circuit, it offers opposition only to *alternating* currents or to *pulsating* direct currents. This singular behavior of reactive opposition is the result of its opposing any *change in amplitude* of a current rather than simply opposing the flow of current. Since an alternating current is continually varying in amplitude, the net result in an a-c inductive circuit is opposition to *total* flow of current. In a theoretical circuit containing only inductance, Ohm's law could be modified for alternating current as follows:

$$E = IX_L, \tag{4}$$

where $I$ and $E$ are effective values and $X_L$ is expressed in ohms. A circuit, therefore, that may present many thousand ohms of inductive reactance to an alternating current will present zero reactance to a uniform smooth direct current. The direct current will have only the normal resistance of the copper wires of the inductive circuit to overcome.

This characteristic is utilized in the wide application of inductances as *choke coils.* A choke coil, or, simply, a choke, derives its name from its tendency to choke out alternating currents to which it offers high reactance. Thus, chokes are used in many places in radio circuits where it is desired to prevent the flow of alternating currents without materially retarding the flow of direct currents. For such an application, a coil is used that has a very high reactance at the frequency of the undesired alternating current. The wire of which the coil is wound is made sufficiently heavy so that the resistance it offers to direct currents can be neglected.

Choke coils are often used where it is desired to discriminate between alternating currents of different frequencies. For example, a coil wound of a few turns of wire with an air core will have so little reactance at audio frequencies as to be negligible. Nevertheless, the reactance of this same coil at radio frequencies may be many thousands of ohms. Such a coil would serve admirably as an r-f choke in any circuit where it is desired to pass audio frequencies with a minimum of attentuation while offering high opposition to radio frequencies.

**Impedance.** The total opposition to an alternating current offered by the combination of inductive reactance and resistance is called **impedance.** In Fig. 69, the resistance of the circuit is the resistive emf divided by the current; the reactance is the inductive emf divided by the current; and the impedance is the total emf divided by the current. The impedance, represented by the letter $Z$, is mathematically equal to the square root of the sum of the resistance squared and the reactance squared. This relation is the result of the resistive and reactive emfs being in quadrature, or 90° apart in phase. The subject of impedance is discussed further in Chap. IX.

## *MUTUAL INDUCTANCE*

Often it is desirable and necessary to utilize two or more inductances in a circuit with the same alternating current flowing through each of the coils. If the coils are physically far enough apart to prevent interaction of their respective magnetic fields, the total inductance of such a series circuit may be computed by simply adding the values of individual inductance. If they should be in close proximity to each other, however, the problem of calculating the effect of the interacting fields upon the total inductance arises. Thus, if two coils connected in series are in close physical proximity to each other and properly oriented one to the other, many of the magnetic lines of force of the first coil are bound to cut the wires of the second coil; and, conversely, a considerable number of the magnetic lines of force of the second coil are bound to cut the turns of the first coil.

The additional inductive effect created in a circuit by the interaction of the magnetic fields of the two coils is called **mutual inductance,** since it is the result of the mutual interaction of the fields of the individual

inductances. It is plain that mutual inductance is dependent upon the values of the individual inductances, since the intensity of the magnetic fields is a function of these values and also upon the degree of coupling, since obviously the relative positions and proximity of the two coils governs the extent of the interaction. The exact relationship is expressed mathematically as follows:

$$M = k\sqrt{L_1 L_2}, \tag{5}$$

where $M$ = mutual inductance in henrys;
$k$ = coefficient of coupling;
$L_1$ = inductance in henrys of one coil;
$L_2$ = inductance in henrys of the other coil.

If the coupling between two coils is perfect, that is if *all* the lines of force of one coil link *all* the turns of the other coil and vice versa, the

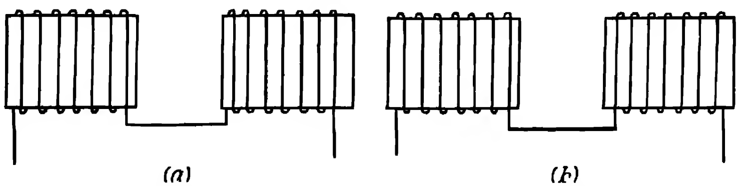


FIG. 70. Relation of windings in series inductance circuits. (a) Series aiding circuit (b) Series opposing circuit

coefficient of coupling is said to be unity, or 1. If some of the lines of force of one coil do *not* link the other and such is *always* the case the coefficient of coupling will be less than unity. It will be seen, therefore, that in actual practice the coefficient of coupling will always take some value between 0 and 1.

When two coils are so connected that their magnetic fields are of the same polarity and are therefore additive, they are said to be connected **series-aiding.** When they are connected so that their magnetic fields are of opposite polarity and therefore buck each other, they are said to be connected **series-opposing.** Hence, it will be noted that the connections of the coils in a series circuit must be considered when computing the total inductance. The value of total inductance for each case is expressed by the following formulas:

$$L_a = L_1 + L_2 + 2M, \tag{6}$$

$$L_o = L_1 + L_2 - 2M, \tag{7}$$

where $L_a$ = total inductance in henrys in series-aiding circuits;
$L_o$ = total inductance in henrys in series-opposing circuits;
$L_1$ = inductance in henrys of first coil;
$L_2$ = inductance in henrys of second coil;
$M$ = mutual inductance in henrys in each case.

**Figure 70(a) illustrates the connection of two coils in a series-aiding circuit. A series-opposing circuit is shown in Fig. 70(b).**

**Problem.** A 100-$\mu$h and a 200-$\mu$h coil are connected series-aiding. The mutual inductance is 50 $\mu$h. What is the total inductance? The coefficient of coupling?

**Solution.** The total inductance, since the circuit is series-aiding, is found by substituting in Formula (6).

$$L_a = L_1 + L_2 + 2M. \tag{6}$$

Reducing the given values to henrys,

$$L_1 = 100\ \mu\text{h} = 0.0001\ \text{h},$$
$$L_2 = 200\ \mu\text{h} = 0.0002\ \text{h},$$
$$M = 50\ \mu\text{h} = 0.00005\ \text{h}.$$

Substituting,

$$L_a = 0.0001 + 0.0002 + (2 \cdot 0.00005) \tag{8}$$

Multiplying,

$$L_a = 0.0001 + 0.0002 + 0.0001 \tag{9}$$

Adding,

$$L_a = 0.0004\ \text{h},$$

or

$$L_a = 400\ \mu\text{h} \tag{10}$$

In order to find the coefficient of coupling, Eq. (5) is utilized. Thus,

$$M = k\sqrt{L_1L_2} \tag{5}$$

Dividing both sides by

$$\sqrt{L_1L_2},$$

$$k = \frac{M}{\sqrt{L_1L_2}} \tag{11}$$

Substituting,

$$k = \frac{0.00005}{\sqrt{0.0001 \cdot 0.0002}} \tag{12}$$

Converting to exponents,

$$k = \frac{5 \cdot 10^{-5}}{\sqrt{2 \cdot 10^{-8}}}. \tag{13}$$

Extracting the root,

$$k = \frac{5 \cdot 10^{-5}}{1.4 \cdot 10^{-4}} \tag{14}$$

Canceling

$$k = \frac{5 \cdot 10^{-1}}{1.4}, \tag{15}$$

and

$$k = \frac{0.5}{1.4}, \tag{16}$$

or

$$k = 0.35 \quad \text{coefficient of coupling.} \tag{17}$$

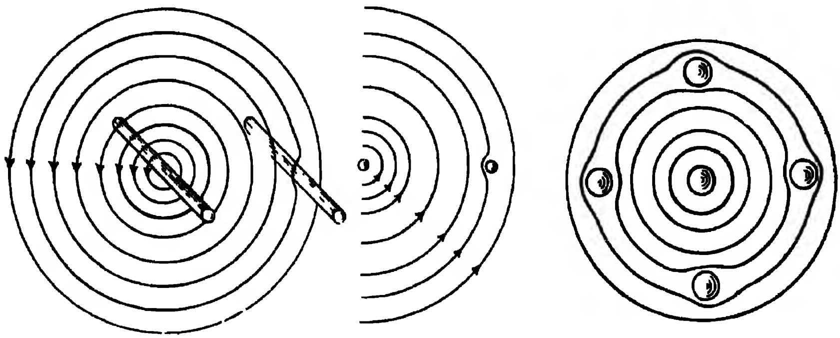
The effects of inductances in parallel with mutual interaction are in general quite complex to calculate. For the most part such calculations are unnecessary, since nothing is gained in practical circuits by using two inductances in parallel when one smaller inductance can achieve the desired result. The inefficiency of connecting two inductances with opposing fields in parallel to *decrease* the total inductance of a circuit is obvious. The mutual inductance of parallel conductors such as are encountered in transmission lines is taken up in detail in the chapter on antennas.

The term **induction** is used when speaking of the inductive effects between separate circuits. Since the induction of external currents is an undesirable condition in most applications, every means must be utilized to keep mutual induction at a minimum. It is apparent that the primary precaution necessary to accomplish this end is to keep the circuit physically as far removed from the fields of other circuits as possible. It is very often impossible to arrange this efficiently, however, and additional steps are therefore necessary to decrease mutual induction. In the comparatively cramped space of radio receivers and transmitters, it is often necessary to mount certain coils relatively close to one another. In general, to minimize the effects of self-induction, such coils are mounted, so far as possible, so that the magnetic field of one coil has a minimum effect on the other. This arrangement is accomplished by making certain that adjacent coils are *never* mounted with their axes in line. Whenever possible, coils should be mounted so that their axes are at right angles to each other. With this arrangement, the relative position of one coil is such that the magnetic lines of force passing through it from the adjacent coil are at a minimum. The field of a given coil can also be confined to a small area about the coil by means of shielding. If a metal shield is so placed that it entirely surrounds a coil, the energy in the alternating field of the coil which strikes this shield is consumed in inducing currents in the shield. The field is therefore effectively stopped at the shield. Aluminum and copper are commonly used for shield construction. Care must be exercised so that a shield is not placed too close to the coil that it is shielding; otherwise, the result would be too great an absorption of energy from the coil, which might materially affect normal circuit operation. It should be remembered that the effective inductance of a shielded coil is less than the inductance of the coil without the shield.

## THE TRANSFORMER

One of the most useful applications of electromagnetic induction is the instrument known as the **transformer.** A transformer is used to transform electric power at one voltage or current value to power at another value of voltage and current.

**Fundamental Principle of the Transformer.** In an earlier part of this chapter, is was seen how the inductance of a conductor could be greatly increased by winding the conductor in the form of a coil. This increase, it was observed, was due to the greater amplitude of the back emf caused by the action of the field of one turn of a coil, generating a back emf not only in itself but also in all the other turns of wire adjacent to it. If a single piece of wire across which an alternating emf is impressed is placed in proximity to another piece of wire *not* connected to the same circuit, an emf will similarly be induced in the second wire (see Fig. 71(a)). If several additional pieces of wire were also placed in the field of the original wire, as in Fig. 71(b), so that practically all the lines of force



(a) (b)

FIG 71 Mutual Inductance.

surround all the wires, it is plain that the same emf would be induced in all wires, since they are all being acted upon by the same field. The original conductor through which alternating current is flowing is called the **primary conductor,** since it is the original source of the magnetic fields. The other conductors in which an emf is being induced are called **secondary conductors.** In Fig. 71(b) there are four secondary conductors. If the four secondaries are connected in series, the emfs across each of them will accumulate. Since all the emfs are equal and since there are four such emfs in the secondary thus formed, the secondary voltage will be four times the primary voltage.

In practical transformers the conductors of the primary and secondary are wound in the form of coils. All possible means are utilized to make the efficiency as high as possible. A high-grade iron core is used to wind the coils about in order to achieve maximum permeability. The secondary voltage, as explained above, is a function of the number of turns of wire in the primary and secondary coils or windings. Specifically, the secondary voltage of any transformer is equal to the primary voltage multiplied by a factor which equals the ratio of secondary turns to primary turns.

Expressed mathematically,

$$\frac{n_p}{n_s} = \frac{E_p}{E_s}, \tag{18}$$

where $n_p$ = number of turns in primary winding;
$n_s$ = number of turns in secondary winding;
$E_p$ = primary voltage;
$E_s$ = secondary voltage.

The ratio $n_s/n_p$ is called the **turns ratio,** since it represents the ratio of secondary turns to primary turns, and is represented in mathematical problems by the letter $N$. Therefore,

$$N = \frac{n_s}{n_p} = \frac{E_s}{E_p}. \tag{19}$$

Transformers are classified as **step-up** or **step-down** transformers according to whether voltage is being stepped up or stepped down by them. In both types of transformer the primary voltage is *multiplied* by the turns ratio $N$ in order to obtain secondary voltage.

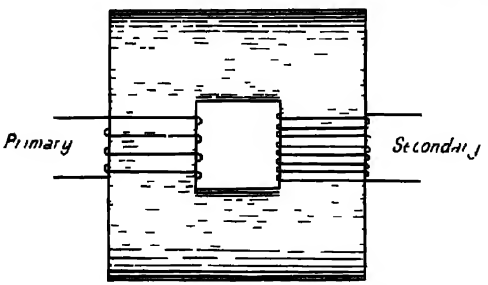


FIG. 72. An elementary step-up transformer of the closed-core type. In this case the step-up ratio is 2 to 1.

It might at first appear that in some unaccountable fashion a transformer is a creator of electric energy, since more voltage is obtained from the output side than is applied to the input side. What is gained on the voltage side, however, is compensated for on the current side. The drop in current in the secondary of a transformer is inversely proportional to the increase in voltage. Thus, in a 100 per cent efficient step-up transformer having a turns ratio of 10, the secondary voltage will be 10 times the primary voltage. The secondary current, however, will be one tenth of the primary current. The power in both windings, neglecting power factor, will be the product of the voltage and current in the winding. The product of the primary voltage and current will equal the product of the secondary voltage and current. No power, therefore, is created within the transformer itself. As a matter of fact, since no transformer is perfect, the power present in the secondary will in actual practice always be slightly less than the primary power due to losses in the transformer. The current ratio in transformers may be expressed by the formula

$$N = \frac{I_p}{I_s}. \tag{20}$$

Combining Eq. (20) with Eq. (19), we may state that

$$N = \frac{n_s}{n_p} = \frac{E_s}{E_p} = \frac{I_p}{I_s}. \tag{21}$$

In a-c networks, the total opposition of all kinds offered to the flow of current is called impedance, as stated earlier. If effective values of voltage and current are used, Ohm's law may be applied to a-c circuits by simply replacing the $R$, which represents the total opposition in d-c circuits, by $Z$, which represents the total opposition to the flow of alternating currents. Thus, for a-c circuits, Ohm's law becomes

$$E = IZ, \tag{22}$$

$$I = \frac{E}{Z}, \tag{23}$$

$$Z = \frac{E}{I}. \tag{24}$$

Assume a step-up transformer with a turns ratio $N$ and with its primary connected to a source of alternating current and its secondary connected to an impedance $Z_s$. By Eq. (21) the secondary voltage would be

$$E_s = NE_p. \tag{25}$$

Similarly, by Eq. (21)

$$I_s = \frac{I_p}{N}. \tag{26}$$

By Ohm's law (Eq. (24)) the apparent primary impedance is

$$Z_p = \frac{E_p}{I_p}. \tag{27}$$

Substituting the equivalent values of $E_s$ and $I_s$ from Eqs. (25) and (26), the secondary impedance becomes

$$Z_s = \frac{NE_p}{\frac{I_p}{N}} = N^2 \cdot \frac{E_p}{I_p}. \tag{28}$$

But, from Eq. (27),

$$\frac{E_p}{I_p} = Z_p. \tag{29}$$

Therefore, substituting in Eq. (28),

$$Z_s = N^2 Z_p. \tag{30}$$

Rearranging,

$$\frac{Z_s}{Z_p} = N^2. \tag{31}$$

The ratio of the secondary to the apparent primary impedance in a transformer is therefore equal to the square of the turns ratio.

**Transformer Losses.** All the above formulas are strictly true only when the transformer is 100 per cent efficient. In actual practice no transformer is perfectly efficient. The efficiency in present day well

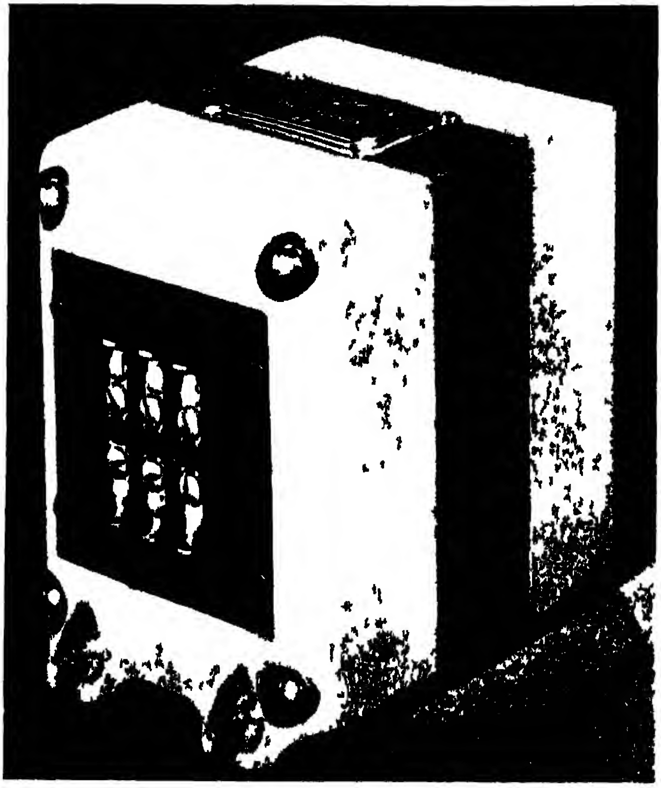


FIG. 71. A typical broadcast transmitter power transformer. (*Courtesy of Thordarson Electric Manufacturing Co.*)

constructed transformers, however, is so great (95 to 98 per cent) that the formulas may be applied with sufficient accuracy for technical work.

The efficiency of a transformer, in common with any other work performing apparatus, is represented by the ratio of output power to input power. Expressed as a formula,

$$\text{efficiency} = \frac{\text{output power}}{\text{input power}} \qquad (32)$$

Maximum efficiency in Eq. (32) would be indicated by unity. This can be converted to percentage by multiplying by 100.

**Problem.** The output of a certain transformer is 50 w. Core losses in the transformer amount to 0.32 w and the copper loss 0.52 w. What is the efficiency?

**Solution.** The input of such a transformer must, of necessity, be equal to the output plus the losses. Substituting in Eq. (32),

$$\text{efficiency} = \frac{50}{50 + 0.32 + 0.52} = \frac{50}{50.84} = 0.983, \tag{33}$$

or

$$\text{efficiency} = 0.983 \cdot 100 = 98.3 \text{ per cent.} \tag{34}$$

If the efficiency of a transformer is known, the voltage and current output can be computed for a transformer of known turns ratio when a given voltage and current are impressed upon the primary.

**Problem.** A step-up transformer with a turns ratio of 5 has an efficiency of 98 per cent. What are the values of secondary voltage and current when 100 v at 10 amp is supplied to the primary?

**Solution.** $100 \cdot 10 = 1{,}000$ w in primary,* that is

$$P = EI \tag{35}$$

$$\frac{\text{output in watts}}{\text{input in watts}} = \text{efficiency} \tag{32}$$

Substituting,

$$\frac{\text{output}}{1{,}000} = 0.98. \tag{36}$$

Therefore,

$$\text{secondary output} = 980 \text{ w}, \tag{37}$$

and

$$\text{apparent impedance in primary} = \frac{100}{10} = 10 \text{ ohms}. \tag{38}$$

By Eq. (31),

$$\text{secondary impedance} = N^2 \cdot 10 = 250 \text{ ohms} \tag{39}$$

By applying the power equation (neglecting power factor), we get

$$P = I^2 Z. \tag{40}$$

Rearranging,

$$I = \sqrt{\frac{P}{Z}}, \tag{41}$$

and

$$I = \sqrt{\frac{980}{250}} = 1.98 \text{ amp} \tag{42}$$

From the power equation,

$$E = \frac{P}{I}. \tag{43}$$

* Actually, the power in an a-c circuit equals $E \cdot I \cdot$ power factor. It can be shown, however, that in any transformer the power factor of the primary circuit very nearly equals the secondary circuit power factor. In problems of this type, therefore, power factor can be neglected. This is discussed further in Chapter IX.

**Substituting,**

$$E = \frac{980}{1.98} = 494 \text{ v.} \tag{44}$$

The output values are therefore 494 v and 1.98 amp.

The losses in transformers have been classified into three general groups, namely, copper losses, iron losses, and leakage losses.

**Copper loss** in a transformer is the loss due to the actual resistance of the wire comprising the coils. The power absorbed by this loss is equal to $I^2R$ in common with resistance losses in other types of circuits. For this reason, it is often called the $I^2R$ loss in transformers. The power expended in overcoming the resistance of the copper wires is dissipated in heating the wires and contributes nothing to the useful function of the unit as a transformer.

**Iron losses,** often called "core losses," are of two kinds: *eddy-current* loss and *hysteresis*. Since the iron that comprises the core of a transformer is itself a conductor, it will have an emf induced in it just as the secondary winding does. The resultant currents that flow in the core from this cause are called **eddy currents.** Eddy currents dissipate their energy in heating the core. **Hysteresis** is the name applied to the energy expended in the iron core of a transformer by reason of molecular friction. Since each individual molecule is in fact a minute particle of iron, each has a north and south pole just as the entire core does. The constant reversals of alternating current flowing through the transformer windings cause the polarity of the individual molecules to change continually. In addition to their normal motion, the molecules are therefore subject to additional movements as a result of the attractive and repellant forces of adjacent magnetized molecules. The resultant increase in molecular agitation causes countless millions of collisions among molecules. The energy expended in the form of molecular friction is drawn from the transformer input source and represents a loss to the circuit.

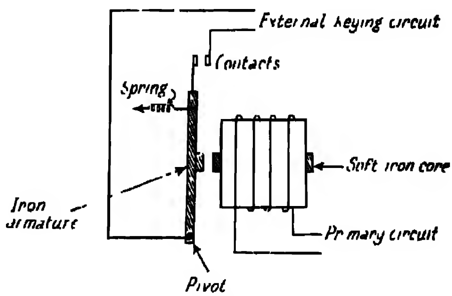
The **leakage loss** in a transformer is a function of the coefficient of coupling between primary and secondary and is the loss experienced because of the lines of force from the primary that never reach the secondary. Since these lines of force return to the primary, however, leakage loss does not constitute a power loss. In modern transformers, the coefficient of coupling approaches unity and leakage loss is very small.

## *THE RELAY*

The principle of electromagnetic inductance is utilized to great advantage in the instrument known as the **relay,** which, basically, consists of a large number of turns of copper wire wound about an iron core. A soft-iron armature is mounted on a hinge or pivot in such a manner that when the iron core is magnetized by current flowing through the coil it

draws the armature to it. When the current is turned off the core loses its magnetism, and the armature is drawn back to its original position by the action of a spring The movement of the armature is utilized to close or open contacts in external circuits. A diagram of an elementary relay is shown in Fig. 74.

A coil wound in the manner of the relay coil in Fig. 74 is called a **solenoid.** The strength of the magnetic poles induced by the solenoid is proportional to the product of the number of turns of wire in the coil and the current flowing through the coil as long as the iron core is not saturated This product is designated as the **ampere-turns** of the solenoid. All other conditions being fixed the pull exerted on the relay armature is proportional to the square of the ampere turns



Elementary keying relay

There are many different types of relays utilized in radio work: overload relays, under load relays, keying relays, switching relays time delay relays, and so on. They will be discussed in detail as they are encountered throughout the text Generally speaking all types of relays operate on the same fundamental principle with various modifications.

## *QUESTIONS AND PROBLEMS**

**1.** What is the unit of inductance?

**2.** What is the inductive reactance of a 30-h choke coil at 100 c?

**3.** Why may a transformer not be used with direct current?

**4.** When two coils, of equal inductance, are connected in series with unity coefficient of coupling and their fields in phase, what is the total inductance of the two coils?

**5.** If a transformer having a turns ratio of 10 to 1 is working into a load impedance of 2,000 ohms and out of a circuit having an impedance of 15 ohms, what value of resistance may be connected across the load to effect an impedance match?

**6.** Define the following terms. hysteresis, permeability, eddy currents.

**7.** What is the meaning of ampere-turns?

**8.** What factors determine the efficiency of a power transformer?

**9.** What factors determine the ratios of primary and secondary currents in a power transformer?

**10.** What is the total inductance of two coils, connected in series, but without any mutual coupling?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

Chapter VIII

# CAPACITANCE

In an earlier chapter the fundamental conception of positively and negatively charged bodies was discussed. It was seen that a positively charged body is one that has a deficiency of electrons, thus making the prevailing charge of the atom positive. Similarly it was noted that a negatively charged body is one having an excess, or surplus, of electrons, thus making the prevailing atomic charge negative. When two such oppositely charged bodies are connected by means of a conductor, a current is caused to flow through the conductor because of the mutual attraction of the oppositely charged atoms. When current flows in such a circuit, it is said that the current flow has been caused by a *difference of potential* between the two bodies. Potential, in this sense, refers to the potential energy per unit charge of a charged body. When a small amount of charge flows from one body to another, the energy expended is equal to the product of the potential difference and the quantity of charge.

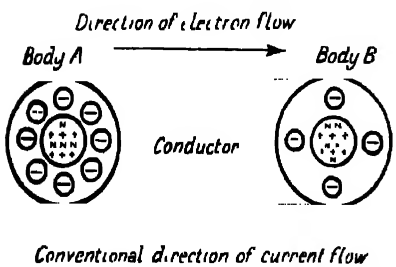



FIG. 75 Difference of potential between charged bodies.

A difference of potential between bodies can exist even though the bodies in question are not oppositely charged. Thus, in Fig. 75, body $A$ is a negatively charged body because of the excess number of electrons contained in it. Body $B$ is a normal body in perfect electrical balance, that is, the negative charges of its electrons are exactly equal to the positive charges of its protons Such a body is neither positive nor negative and is called a neutral body. Nevertheless, a difference of electrical charges, that is, a difference of potential *does* exist between bodies $A$ and $B$. Consequently, if bodies $A$ and $B$ are connected by a conductor, current will flow through the conductor. The convention has always been that current flows away from the body with the more positive potential toward the body with the more negative potential. Here the body $B$ has the more positive charge, and the body $A$ the more negative charge, so a current is said to flow from $B$ to $A$, although in actual fact negative electricity (electrons) flows from $A$ to $B$. In computing and

designing circuits, it is best to use the standard convention for current flow rather than to consider the electron flow.

Since body *B* has fewer negative charges than body *A*, body *B* may be said to be *less negative* than body *A*. Body *B*, therefore, may be considered *positive with respect to body A*. This statement does not refer to the *absolute* charge of body *B* but is simply a statement of the relation of the charges between these two bodies.

This point can perhaps be more clearly understood by reference to a simple series resistance circuit such as that of Fig. 76. If the three 10-ohm resistors of Fig. 76 are connected across a 90-v source of direct current, according to Ohm's law the current through the circuit is 3 amp. Since it is a series circuit, 3 amp of current will flow through all three of the resistors. Applying Ohm's law to each of the resistors in turn, the *voltage drop* or *difference of potential* across each of the resistors is found to be 30 volts. According to the battery polarity shown in the illustration, point *D* is positive with respect to any other point in the circuit. Point *C* is negative with respect to *D* by 30 v. Nevertheless, point *C* is *positive* with respect to point *B* by 30 v. It is also positive with respect to point *A* by 60 v. Thus it can be seen that the *relative charge* of any point in a circuit is independent of the atomic structure, but it is a function of the charges of all other points in that circuit.

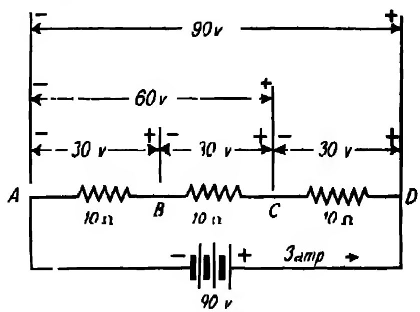


FIG. 76. Relative polarities of different points in a series circuit.

The quantity of electrical charge on a conductor depends on the potential of the conductor and also on the size, shape, and location of the conductor with respect to other material. The ability of a conductor to store a charge under a given potential difference is called its **capacitance.**

## EFFECTS OF CAPACITANCE

**Capacitance in A-C Circuits.** If two linear conductors are connected to a source of alternating voltage, as shown in Fig. 77(a), they will alternately charge and discharge as the voltage at the source rises and falls. At any particular instant, the conductor that is connected to the positive side of the alternator will have a deficiency of electrons, or a positive charge, and the conductor that is connected to the negative side will have an excess of electrons, or a negative charge. The charges are such that the potential difference between the charges is equal to the potential difference at the source of voltage. When the voltage source

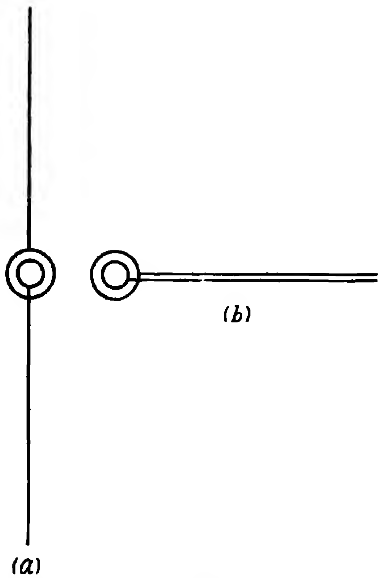



Fig. 77. Capacitive effect of conductors.

reverses in direction, the charges reverse, and the electrons flow from one conductor through the source to the other conductor. This results in a current flow that may be measured with a sensitive ammeter if the conductors are long enough.

It can therefore be seen that current may be drawn from an alternating source without actually connecting the terminals together. The charge will be increased, and, hence, the current increased, if the conductors are brought closer together, as shown in Fig. 77(b). Since unlike charges attract each other, the proximity of a charge on one conductor will attract *more* charge of opposite sign on the other conductor, even though the potential difference is the same as for Fig. 77(a). Thus, the capacitance between conductors is increased if they are moved closer together. The capacitance is still further increased if the conductors are formed in the shape of flat plates and brought near each other with the faces parallel. A device consisting of two or more such conductors is called a **capacitor.***

Figure 78(a) illustrates a capacitor circuit connected to a source of alternating emf. Figure 79 is a graph of a sine wave of the alternating emf impressed upon the circuit of Fig. 78(a) with the resultant current flow. At zero phase angle in Fig. 79 (the start of the voltage curve), the exact moment the voltage starts

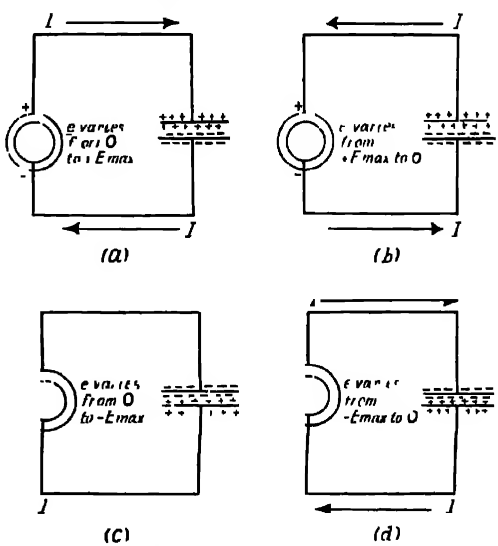



Fig. 78. Current flow in a capacitive circuit for a single cycle of impressed alternating emf.

* The term "capacitor" is now more generally used than "condenser," which was formerly in general use.

to rise, a difference of potential exists between the positive terminal of the alternator, Fig. 78(a), and the upper plate of the capacitor, because the neutral atoms of the upper capacitor plate are actually negative with respect to the generator positive terminal. The latter is then made deficient in electrons by the action of the generator. Current, by convention, flows to the upper capacitor plate from the positive alternator terminal as indicated by the arrow.

In a similar manner, current will flow from the lower capacitor plate to the negative generator terminal, and the charge on the lower plate will be equal and opposite to the charge on the upper plate. Since the total charge is always proportional to the potential difference imposed by the generator, the charge will change as the generator voltage changes.

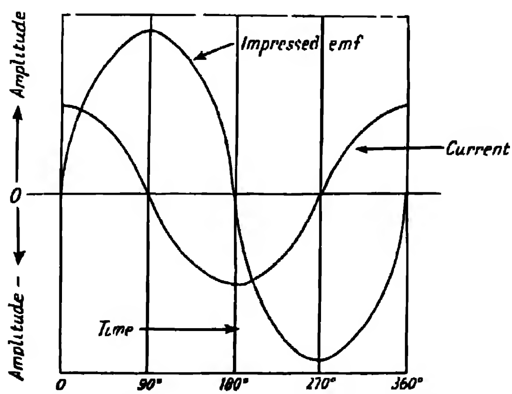



FIG. 79 Voltage and current displacement for the circuit of Fig 78

Since, however, a change in charge can be brought about only by a flow of current, a current flows when the applied voltage is *changing*, and no current flows when the voltage is constant. The current flow is a maximum when the voltage is zero, because the rate of change of a sinusoidally varying voltage is greatest at this moment. This moment occurs at the beginning of the first quarter cycle (from 0 to 90 ) of Fig. 79 and is represented in Fig. 78(a). For the next quarter cycle (from 90 to 180 ) the applied emf is decreasing, and the direction of current is reversed, as shown in Fig. 78(b). During the next quarter cycle, the emf continues to decrease until the maximum negative value is reached, and the current, as shown by Fig. 78(c), is in the same direction as in Fig. 78(b). During the last quarter cycle, the emf is rising again, and the current again flows in the original direction.

It is apparent that the current in such a purely capacitive circuit *leads* the voltage by 90°. Because all circuits possess resistance, this angle of lead would never be attained in actual practice. Since 90° lead

represents the condition in a theoretical purely capacitive circuit, the lead angle in an actual capacitive circuit will always take some value between 0 and 90°.

Although current in the sense of ordinary conduction current does not actually flow between the plates of the capacitor, conduction current *does* flow through every other part of the circuit, as it moves to and from the capacitor plates.* Except for the phase displacement of 90°, this current through the circuit is identical with that which would flow were the circuit closed. For example, if two incandescent lamps were connected in series with such a circuit as that shown in Fig. 80, the lamps could be made to light to full brilliancy because of the current flowing through them, just as they could in a closed circuit. For convenience in referring

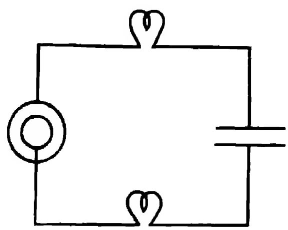


FIG. 80. Two incandescent lamps connected in series

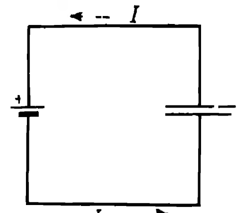



FIG. 81. Capacitance in a d-c circuit

to such circuits, therefore, it is often said that alternating currents will flow *through* a capacitor (see footnote below).

Since the phenomenon of capacitance depends upon the charges on the plates of a capacitor, capacitance is a function of the *size* of the plates. The larger the area of the plates, the greater the charge on the plates of a capacitor for a given potential and, consequently, the greater the capacitance of the capacitor. It may be said, therefore, that in general *the capacitance of a capacitor is directly proportional to the area of the plates.*

Inasmuch as the potential difference between given charges increases as the distance between them decreases, it is apparent that the capacitance of a capacitor is also a function of the distance between the plates. Specifically, it can be said that *the capacitance of a capacitor is inversely proportional to the distance between the plates.*

* Whenever the space intervening between the plates of a capacitor (or between an antenna and ground) is subjected to a time-varying potential difference, this region is subjected to a time-varying electric field. A time-varying electric field, even though it be in a vacuum where no charge carriers are present, is the *equivalent* of an electric current. This equivalent electric current is called Maxwell's displacement current to distinguish it from ordinary conduction current, which takes place in material conductors. Displacement currents are of considerable importance in understanding advanced subjects like radiation from antennas and propagation of electric energy through hollow wave guides. In the case of an ordinary capacitor, the displacement current that flows in the region between the plates is equal in magnitude and phase to the conduction current that flows in the remainder of the circuit. Thus the current is rendered continuous in the series circuit.

The unit of capacitance is the **farad,** named after Michael Faraday, the English scientist. *A capacitor is said to have a capacitance of one farad when a difference of potential of one volt between the plates of the condenser will create a charge of one coulomb on the plates.* A farad is too large a unit for convenient practical use, so the microfarad (one millionth of a farad, abbreviated $\mu f$) is used for radio work as well as the micro-microfarad (one millionth of a microfarad, abbreviated $\mu\mu f$).

**Capacitance in D-C Circuits.** If a capacitor were connected to a source of direct current, as shown in Fig. 81, the action in every way would be similar to that during the first quarter cycle with an a-c source. [See Fig. 78(a).] The current flow would be great in amplitude at the moment the circuit is closed, diminishing to zero as the voltage approached the maximum battery voltage. At this point, however, the battery voltage, unlike an alternating voltage, remains constant. There is no change of potential, and the current remains at zero. If incandescent lamps were connected in series with the legs of this circuit, there would be only a momentary flicker of light when the circuit was closed and current flowed for a brief instant. So far as useful applications are concerned, therefore, a capacitive circuit is an open circuit to direct currents. In the phraseology of the art, direct currents cannot pass "through" a capacitor

If the capacitor in Fig. 81 is disconnected from the battery, the atoms of the capacitor plates will be found to retain the unbalanced state caused by the application of battery voltage In other words, the upper plate would have a positive charge, and the lower plate would have a negative charge A capacitor in this state is said to have been "charged." If the capacitor plates are shorted together by means of a piece of wire, the capacitor will "discharge." The electrons rush from negative to positive plate to restore the normal equilibrium of the atoms, which is evidenced by a spark when contact is made.

Theoretically, a perfect capacitor will retain its charge indefinitely. In practice, however, it is found that no capacitor is perfect, since no insulator is perfect. The insulating material between the plates of a capacitor is called the **dielectric.** A charged capacitor loses its charge within a short time (depending upon the capacitance of the capacitor) because of the leakage between the plates through the dielectric.

The amount of charge that a capacitor is capable of assimilating depends upon the capacitance of the capacitor and the voltage impressed across the plates. Expressed mathematically,

$$Q = CE, \tag{1}$$

where $Q$ = charge in coulombs;
$C$ = capacitance in farads;
$E$ = potential across $C$ in volts.

The amount of work done in charging a capacitor depends upon the capacitance of the capacitor and the voltage to which it is charged. This relation is expressed as the formula

$$W = \tfrac{1}{2}CE^2, \tag{2}$$

where $C$ = capacitance of the capacitor in farads;
$E$ = voltage to which the capacitor is charged in volts;
$W$ = work accomplished in joules.

## *CAPACITIVE REACTANCE*

**Capacitor Losses.** A capacitor, like every other electrical instrument, is not 100 per cent efficient; when it is charged, a certain amount of energy is consumed in overcoming the inherent losses of the condenser. The losses incurred in a capacitor are of several kinds. **Dielectric hysteresis** loss is occasioned by the heat produced in the dielectric of a condenser by molecular friction. **Brush discharge** is a form of loss, which will be discussed in detail in Chap. XV. It is particularly noticeable when the edges of capacitor plates are greasy or soiled. **Dielectric absorption** losses occur when a capacitor sets up an electric field between it and other objects in close proximity to it. When capacitors are injudiciously mounted too close to inductances in r f circuits, this loss can assume sizable proportions. Because no dielectric material is a perfect insulator, there is also a loss in capacitors due to current **leakage** between the plates.

Whenever an emf is required to produce a flow of current in a circuit, the circuit is said to offer opposition to the flow of current. The peculiar type of opposition offered by capacitance is called **capacitive reactance.** In common with other forms of opposition to the flow of current, it is expressed in ohms. The ohm of capacitive reactance is similar to the ohm of inductive reactance, inasmuch as both offer opposition to the flow of current in a circuit as well as affect phase displacement. A capacitive circuit, however, tends to cause the current to *lead* the voltage by 90°. An inductive circuit tends to cause the current to lag the voltage by 90°. The displacement effect of a capacitive circuit, therefore, is 180° out of phase with that of an inductive circuit; or, in another way of stating it, the phase displacement effect of capacitive reactance is exactly opposite to inductive reactance.

An examination of Fig. 79 will disclose that when a capacitor is connected to a source of alternating current, maximum current flows in the circuit only at certain periods. These periods are the beginning of each alternation—at 0 and 180° phase angles where the rate of change of emf is greatest. Therefore, the greater the frequency, the greater the amplitude of current, because the voltage is changing faster. Since more current flows as the frequency increases, this is equivalent to saying that

capacitive reactance varies inversely as the frequency. The mathematical formula for capacitive reactance is

$$X_c = \frac{1}{2\pi f C}, \tag{3}$$

where $X_c$ = capacitive reactance in ohms;
$\pi$ = the constant 3.14;
$f$ = frequency in cycles per second;
$C$ = capacitance in farads.

Although throughout this discussion a capacitor has been represented as having only two plates, this assumption has been made simply to facilitate explanation. In actual practice, capacitors are more often

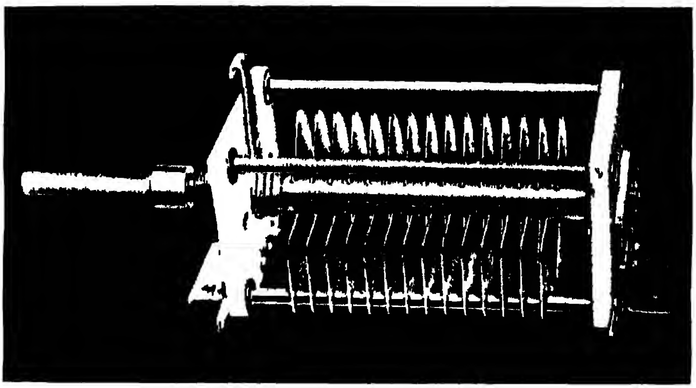


FIG. 82. A typical variable capacitor. *(Courtesy of Hammarlund Manufacturing Co., Inc.)*

constructed of a greater number of plates in order to obtain the greatest possible capacity utilizing the smallest space. Commercial capacitors are often constructed of dozens of tinfoil strips arranged in layers with wax impregnated paper separating the layers. Alternate layers are connected to common terminals. Glass, paper, oil, compressed air, and mica are dielectrics commonly used in commercial capacitors.

A fixed capacitor is one in which the capacitance is a fixed quantity; that is, the number or size of the plates, the nature of the dielectric, and the spacing between plates cannot be varied. A variable capacitor is one that has been purposely constructed so that its capacitance can be changed at will. The change is usually accomplished by mounting one set of plates upon a movable shaft in such a manner that these plates (called **rotors**) can be rotated, or meshed, between the stationary plates, which are called **stators**. Figure 82 illustrates the construction of a typical commercial capacitor of the variable type.

**Problem.** If a capacitor has a capacitance of 10 $\mu$f, what will its reactance be at a frequency of 1 kc?

**Solution.** The given values of the problem are prepared for substitution in Eq. (3) by reducing them to the units required, thus,

$$C = 10\ \mu\text{f} = 0.00001\ \text{f},$$

$$f = 1\ \text{kc} = 1{,}000\ \text{c}.$$

$$X_c = \frac{1}{2\pi fC}, \tag{3}$$

$$X_c = \frac{1}{6.28 \cdot 10^3 \cdot 1 \cdot 10^{-5}}, \tag{4}$$

$$X_c = \frac{1}{0.0628}, \tag{5}$$

$$X_c = 15.9\ \text{ohms}. \tag{6}$$

It has been shown that direct currents cannot flow through a capacitor. High-frequency alternating currents, however, flow in a capacitor circuit almost as well as in a closed circuit. Consequently, one of the most common applications of capacitors in radio circuits is in cases where it is desired to prevent the flow of direct currents while simultaneously permitting alternating currents to pass relatively unimpeded. For such applications, a value of capacitance is chosen which offers a sufficiently low capacitive reactance at the frequency of the alternating current it is desired to pass. At the same time, care must be taken that the voltage rating of the capacitor is sufficiently high to preclude the possibility of puncture or breakdown of the capacitor dielectric For this usage, both the d-c voltage and the a-c voltage of the circuit must be considered, since both are being applied to the capacitor plates.

Capacitors are also used where it is desired to discriminate between alternating currents of different frequencies. This usage is possible because of the fact that the capacitive reactance of a capacitor varies inversely with the frequency. Thus, whereas a given capacitor may offer a very low reactance to a high frequency, the same capacitor may offer a reactance of many thousands of ohms to l-f alternating currents. Capacitors are often utilized in this manner in radio receivers where it is desired to prevent r-f currents from entering the audio system of the receiver without shunting away any of the desired a f currents. For example, a capacitor of 0.005-$\mu$f capacitance is often connected across the detector output circuit of a receiver to ground. Such a capacitor offers very little reactance to r-f currents. and the latter are effectively grounded through the capacitor and thus prevented from entering the following audio stage. The same capacitor, however, offers such a high reactance to a-f currents that the flow of such currents through the capacitor to ground is negligible. Almost all the a-f currents are permitted to enter the following audio stage.

A combination of inductance and capacitance when properly connected in a circuit presents an ideal arrangement for discrimination purposes. Such a combination is called a **filter** and is discussed at length under power supplies in a later chapter.

## *THE ELECTROLYTIC CAPACITOR*

It has been known for some time that the oxides of certain metals possess peculiar electrical characteristics. It has been found that the oxides of tantalum and aluminum possess particularly desirable characteristics for use in capacitors, although the cost of the tantalum has limited its economic usefulness for commercial electrical applications. Consequently, aluminum, being both plentiful and sufficiently economical, has become the most widely used metal for electrical applications.

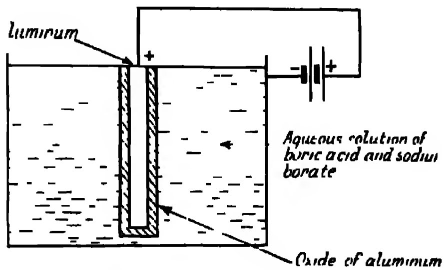
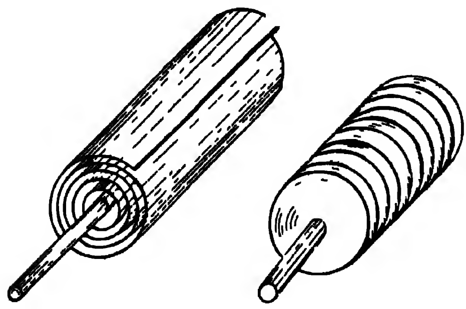



Fig. 83. Forming aluminum oxide.

The oxide of aluminum possesses the remarkable characteristic of permitting the flow of an electric current through it only in one direction when immersed in a bath of conducting fluid. Current will flow quite readily from the aluminum through the aluminum oxide into the electrolyte, which is the technical name for the conducting fluid. If the polarity is reversed, however, it is found that the current flow is reduced to so small an amount as to be considered negligible for most applications. With this reversed polarity, the oxide of aluminum has been found to possess extraordinary insulating qualities, an aluminum oxide coating being capable of withstanding the incredible pressure of *ten million volts per centimeter* without rupture.

An oxide is easily formed on aluminum by electrolytic means. The metal is immersed in an electrolyte usually composed of an aqueous solution of boric acid and sodium borate, as shown in Fig. 83. When an electric current is passed through the solution with the polarity as shown in the diagram, electrolysis occurs. The oxygen that is evolved at the positive pole oxidizes the surface of the aluminum. As soon as the aluminum anode is completely covered with a film of oxide, the oxide offers such a very high resistance to the further passage of current that additional coating of the metal with oxide does not occur. The film of oxide formed over the aluminum by this means is approximately 0.00005 cm in thickness. Despite its extreme thinness, it can, because of its remarkable insulating properties, safely withstand several hundred volts without puncture.

**Wet Electrolytic Capacitors.** Since the capacitance of a condenser is inversely proportional to the distance between the plates, it is apparent that aluminum oxide is an ideal condenser dielectric for some applications. The extreme thinness of such a dielectric makes it possible to construct capacitors of comparatively large capacitance with much smaller physical dimensions than is possible when any other type of dielectric is utilized. This is the primary advantage of the electrolytic capacitor. The use of such a capacitor, however, is limited to d-c applications, inasmuch as the proper polarity must be observed. Electrolytic capacitors find wide application in power-supply filter circuits where pulsating direct currents exist, and they are also used to advantage in by-pass circuits where it is desired to permit the passage of alternating currents while simultaneously blocking direct currents. In the latter application, proper polarity need only be observed with respect to the direct current.



(a) (b)

FIG. 84 Typical electrolytic capacitor anodes (a) Spirally wound wet electrolytic anode (b) Parallel disk type of wet electrolytic anode

Wet electrolytic capacitors are so called because the electrolyte is maintained in liquid form. Commercial wet electrolytic capacitors are constructed of a number of oxidized aluminum plates immersed in a suitable electrolyte. The electrolyte container (usually constructed of aluminum, though *not* oxidized) forms the negative pole of the capacitor. Since the electrolyte is a conductor, it is in series with the negative terminal, or container, thus effectively bringing the negative terminal into actual contact with the oxide film. Commercial capacitors are often constructed with the aluminum anode in the form of a single thin sheet of metal that has been wound in the form of a cylindrical spiral, Fig 84(a). Some manufacturers construct the anode of a number of parallel aluminum disks mounted as closely together as practical with both sides exposed to the electrolyte, as shown in Fig. 84(b).

**Dry Electrolytic Capacitors.** The dry electrolytic capacitor is so designated because the electrolyte contains a very low water content. Such a nonaqueous electrolyte is a relatively poorer conductor than an aqueous electrolyte. Since the negative terminal, or cathode, is in series with the electrolyte, the type of construction utilized in the wet capacitor is not suitable for a dry capacitor. Dry electrolytic capacitors are constructed of alternate layers of cathode and anode capacitor plates interspersed with electrolyte-impregnated separators. The cathode plates are

of pure aluminum. The anode plates are of aluminum coated on both sides with a film of aluminum oxide. The customary form of construction utilizes two separators, a cathode plate, and an anode plate wound in the form of a concentric roll, as shown in Fig. 85. In order to provide complete coverage of the anode-plate dielectric film, the cathode plate is on the outside and completely encircles the winding.

Older-type dry electrolytic capacitors utilize separators made of cotton gauze impregnated with electrolyte. Later-type capacitors employ chemically pure separators constructed of specifically fabricated paper or cellulose impregnated with electrolyte. A number of different electrolytes are utilized for impregnation. Glycoammonium borate compounds are commonly used, but there are hundreds of suitable chemical mixtures.

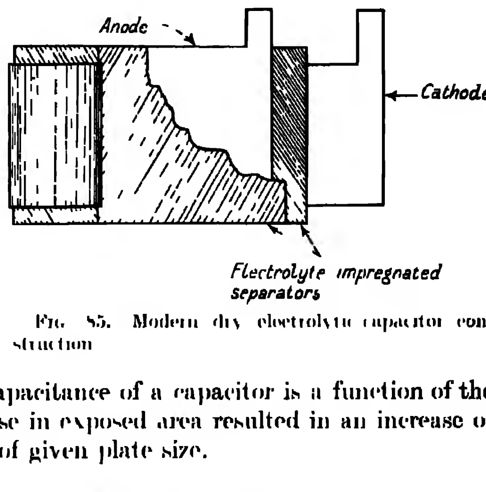



Fig. 85. Modern dry electrolytic capacitor construction

In recent years the capacity of both dry and wet electrolytic capacitors was considerably increased by means of etching. It was found that if the aluminum plates were etched by passing them through a hydrochloric acid solution, the resultant roughening of the plate surfaces greatly increased the effective area of the plates exposed to the electrolyte. Since the capacitance of a capacitor is a function of the area of the plates, the increase in exposed area resulted in an increase of capacitance for a capacitor of given plate size.

## *PARALLEL CAPACITANCE*

The total capacitance of a circuit composed of a number of capacitors connected in parallel is equal to the simple sum of the individual capacitances. Expressed mathematically,

$$C_t = C_1 + C_2 + C_3 + \cdots . \tag{7}$$

## *SERIES CAPACITANCE*

The total capacitance of a circuit composed of a number of capacitors connected in series is a more complex quantity than that of parallel circuits. It is equal to the reciprocal of the sum of the reciprocals of the individual capacitances. This can be more clearly expressed mathematically as

$$C_t = \frac{1}{\dfrac{1}{C_1} + \dfrac{1}{C_2} + \dfrac{1}{C_3} + \cdots}, \tag{8}$$

or, expressed differently, as

$$\frac{1}{C_t} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots. \tag{9}$$

For the special case where there are only two capacitors in series, the following formula applies.

$$C_t = \frac{C_1 C_2}{C_1 + C_2} \tag{10}$$

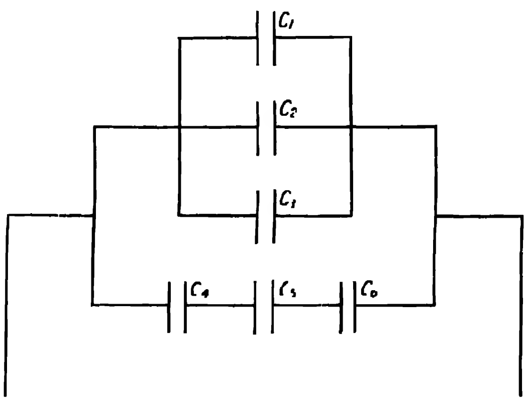



Fig. 86. Capacitors in series parallel.

Another special condition exists when all the capacitors in a series circuit have the same capacitance. The following formula then applies:

$$C_t = \frac{C_s}{N}, \tag{11}$$

where $C_t$ = total capacitance;
$C_s$ = capacitance of any *one* capacitor;
$N$ = number of capacitors.

Formulas (10) and (11) hold true only for the special conditions designated. Formulas (8) and (9) hold true for *any* condition, including the special conditions of Formulas (10) and (11). The latter two formulas are *additional* formulas, which may be used as a matter of preference and simplification.

**Problem.** In the circuit of Fig. 86, $C_1 = 1\ \mu f$, $C_2 = 3\ \mu f$, $C_3 = 5\ \mu f$, $C_4 = 5\ \mu f$, $C_5 = 3\ \mu f$, and $C_6 = 7\ \mu f$. What is the total capacitive reactance of the circuit at a frequency of 1,000 c?

**Solution.** In the parallel circuit comprised of $C_1$, $C_2$, and $C_3$, according to Formula (7), the total capacitance is

$$C_{tp} = C_1 + C_2 + C_3. \tag{12}$$

Substituting,

$$C_{tp} = 1 + 3 + 5 \tag{13}$$

and

$$C_{tp} = 9\ \mu\text{f}. \tag{14}$$

In the series circuit comprising $C_4$, $C_5$, and $C_6$, according to Formula (8), the total capacitance is

$$C_{ts} = \frac{1}{\frac{1}{C_4} + \frac{1}{C_5} + \frac{1}{C_6}}. \tag{15}$$

Substituting,

$$C_{ts} = \frac{1}{\frac{1}{5} + \frac{1}{3} + \frac{1}{7}}, \tag{16}$$

and

$$C_{ts} = \frac{1}{0.676} = 1.48\ \mu\text{f}. \tag{17}$$

In the parallel circuit comprising $C_{tp}$ and $C_{ts}$, according to Formula (7), the total capacitance is

$$C_t = C_{ts} + C_{tp}. \tag{18}$$

Substituting,

$$C_t = 1.48 + 9, \tag{19}$$

and

$$C_t = 10.48\ \mu\text{f}.$$

According to Formula (3), the reactance of the circuit is equal to

$$X_c = \frac{1}{2\pi f C}. \tag{20}$$

Substituting,

$$X_c = \frac{1}{6.28 \cdot 1{,}000 \cdot 0.00001048}, \tag{21}$$

where total $C$ has been reduced to farads. Simplifying,

$$X_c = \frac{1}{6.28 \cdot 10^3 \cdot 1{,}048 \cdot 10^{-8}}, \tag{22}$$

$$X_c = \frac{10^5}{6.28 \cdot 1{,}048}, \tag{23}$$

$$X_c = 15.1 \text{ ohms} \tag{24}$$

It will be noted that there is a certain similarity between the formulas for total capacitance and total resistance in d-c circuits. The form of the formulas is the same except that the series d-c form holds true for *parallel* capacitance circuits, and the parallel d-c form holds true for

*series* capacitance circuits. Of course in capacitance circuits the unit employed is the farad, whereas in resistance circuits the ohm is used.

Fig. 87(a). Low-capacity high-voltage ... *(Courtesy of Cornell-Dubilier Electric Corp.)*

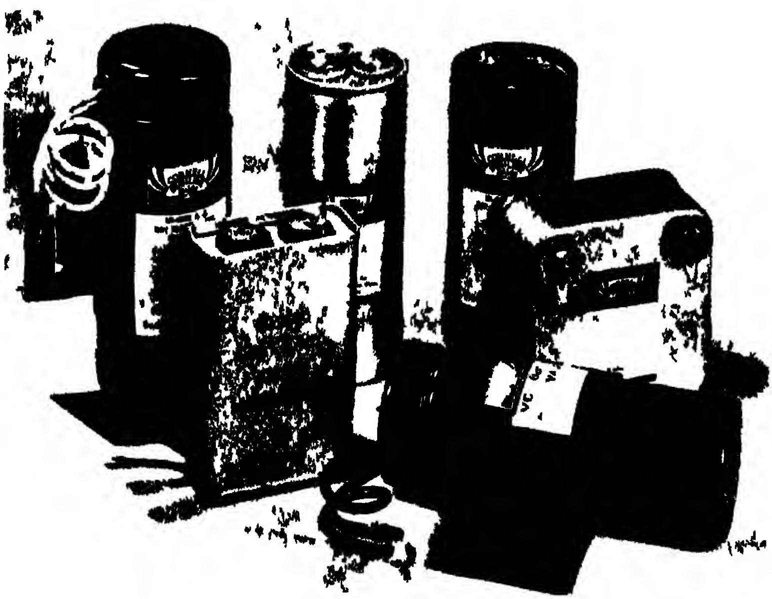


Fig. 87(b). A group of typical wet and dry electrolytic capacitors. *(Courtesy of Cornell-Dubilier Electric Corp.)*

## *QUESTIONS AND PROBLEMS**

**1.** Explain the effect of increasing the number of plates upon the capacitance of a capacitor.

* These questions and problems are taken from the F.C.C. Study Guide for Commercial Radio Operator Examinations.

2. State the formula to determine the capacitive reactance of a capacitor.

3. If capacitors of 5, 3, and 7 $\mu$f are connected in series, what is the total capacitance?

4. The charge in a capacitor is stored in what portion of the capacitor?

Fig. 87(c). Capacitors utilizing oil dielectric. (*Courtesy of Cornell-Dubilier Electric Corp.*)

Fig. 87(d). Typical paper dielectric capacitors. (*Courtesy of Cornell-Dubilier Electric Corp.*)

5. Having available a number of capacitors rated at 400 v and 2 $\mu$f each, how many of these capacitors would be necessary to obtain a combination rated at 1,600 v 1.5 $\mu$f?

6. May two capacitors of 500 v operating voltage, one an electrolytic and the other a paper capacitor, be used successfully in series across a potential of 1,000 v? Explain your answer.

7. How may a filter capacitor be checked for leakage?

8. What is the total reactance when two capacitances of equal value are connected in series?

9. What is the reactance of a capacitor at the frequency of 1,200 kc if its reactance is 300 ohms at 680 kc?

10. When capacitors are connected in series in order that the total operating voltage of the series connection is adequate for the output voltage of a filter system, what is the purpose of placing resistors of high value in shunt with each individual capacitor?

## Chapter IX

# ADVANCED ALTERNATING-CURRENT THEORY

In a-c networks, as in d-c networks, there are two fundamental types of circuits to be dealt with, namely, the series circuit and the parallel circuit. Any complex network, no matter how involved, may be broken down into a number of series or parallel circuits or combinations of both. Unlike d-c networks, however, there is more than merely resistance with which to contend in a-c circuits. The peculiar type of opposition offered to the flow of alternating currents by the effects of inductance and capacitance in the circuit must also be taken into consideration. These effects are found to cause quite different voltage and current relationships from those experienced with direct currents. To determine specifically just what these relationships are in series and in parallel circuits is what comprises the study of a-c theory.

### *SERIES A-C CIRCUITS*

**Current in Series Circuits.** The current in a series a-c circuit is the same in all parts of the circuit. In other words, the current rule for series circuits holds true for both d-c and a-c networks. In the circuit shown in Fig. 88 an inductance, a capacitance, and a resistance are connected in series to an a-c source. If the current flowing through the inductance is represented as $I_L$, the current flowing through the capacitance as $I_C$, and that through the resistance as $I_R$, it can be stated mathematically; thus,

$$I_L = I_C = I_R = I_T, \tag{1}$$

where $I_T$ is the total current or the current flowing through the generator.

**Voltage in Series Circuits.** It was seen in an earlier chapter that inductive and capacitive reactances are opposite in their effects upon alternating current. In the series circuit of Fig. 88, since the current is the same in all parts of the circuit, these opposing effects must of necessity be evidenced in the voltage. It was seen also that the effect of inductive reactance in a circuit is to displace the current from the voltage by 90°. Similarly, it was noted that the effect of capacitive reactance in a circuit is to displace the current from the voltage by 90°. However, the displacements are in opposite directions, inductive circuits causing lagging

currents and capacitive circuits causing leading currents. Since by definition, the current is the same in all parts of a series circuit, the net effect of the reactances in such a circuit is to cause the voltages across the reactances to be displaced with respect to each other. Thus, in Fig. 88, according to Eq. (1),

$$I_L = I_C. \tag{2}$$

Since in an inductive reactance, the current lags the voltage by 90°, the voltage across $L$, or $E_L$, must lead $I_L$ by 90°. Therefore, by Eq. (2), $E_L$ leads $I_C$ by 90°. In the capacitive reactance, the current leads the voltage by 90°. Hence, the voltage across $C$, or $E_C$, must lag $I_C$ by 90°. Since one voltage lags the current by 90° and the other voltage leads the *same* current by 90°, it is apparent that one voltage leads the other voltage by 180°. In other words, $E_L$ and $E_C$ are displaced by 180°. A displacement of 180° makes these voltages at any instant opposite in polarity to each other.

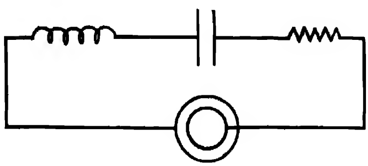


Fig. 88.—Inductance, capacitance, and resistance in series.

In a d-c series circuit, the total voltage across the circuit is the simple sum of the voltages across the component parts of the circuit. In an a-c series circuit, the component voltages are also added to obtain the total voltage. In the latter, however, the component voltages have different phase displacements as well as absolute values and must be considered as vector quantities. Therefore, *the total voltage across an a-c series circuit is the vector sum of the component voltages.*

This relation is represented graphically in Fig. 89. If the current in the circuit of Fig. 88 is represented on the $X$ axis, the voltage across the resistance, or $E_R$, is also represented on the $X$ axis, since it is in phase with the current. This voltage is represented in Fig. 89 by the vector $OA$. Since the voltage across the inductance leads the current by 90°, $E_L$ is represented on the graph by the vector $OB$, which differs in direction from vector $OA$ by 90°. Since the capacitive voltage lags the current by 90°, $E_C$ is represented on the graph by the vector $OC$ which differs in direction from vector $OA$ by 90° in the opposite direction. Inasmuch as $E_L$ and $E_C$ are opposite to each other, they can be added algebraically to obtain the total effective reactive voltage $E_X$, shown in Fig. 89 by the vector $OD$. For purposes of illustration, $E_L$ was taken to be larger in value than $E_C$, and the vector $OD$ appears on the $E_L$ side of the $X$ axis. The total, or resultant, voltage across the entire circuit is obviously a function of $E_X$ and $E_R$ and is found by constructing a parallelogram of forces having sides $OA$ and $OD$, shown in Fig. 89 by parallelogram $ODFA$. The vector $OF$ is the resultant, or total, voltage; that is, the

resultant voltage will have a numerical value equal to the length of the vector $OF$. The phase displacement (phase angle) of the total circuit voltage with respect to the current will be represented by the angle $\theta$.

$E_L$ and $E_C$ will always be 180° out of phase. Consequently, the parallelogram will always be a rectangle (or square), regardless of voltage values. $E_T$, therefore, can easily be computed by means of the Pythagorean theorem (Chap. II), since triangle $OFA$ will always be a right triangle.

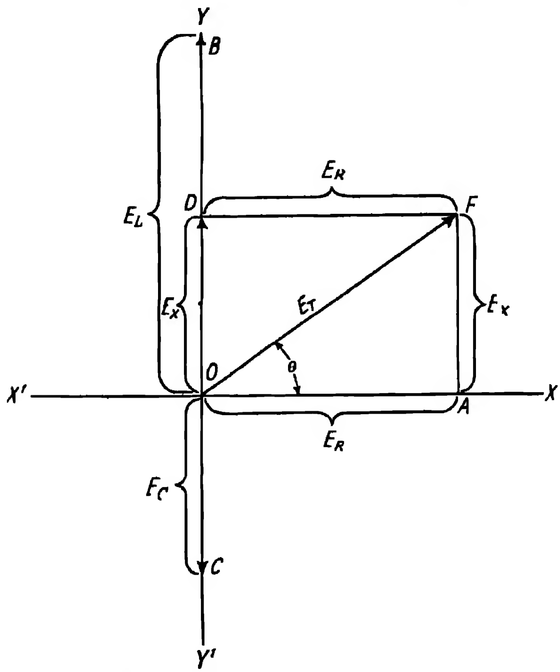



FIG. 89. Voltage vector diagram for a series circuit containing inductance, capacitance, and resistance.

According to the Pythagorean theorem, in any right triangle, the square of the hypotenuse is equal to the sum of the squares of the remaining two sides. Hence,

$$E_T = \sqrt{E_R^2 + E_X^2}. \tag{3}$$

Inasmuch as $E_X$ is the algebraic sum of $E_L$ and $E_C$, these original voltage values may be substituted in Eq. (3), giving the final formula

$$E_T = \sqrt{E_R^2 + (E_L - E_C)^2}. \tag{4}$$

Equation (4) is the formula, therefore, for total voltage in series a-c circuits, and substitution can be made directly in it for the solution of problems. A vector diagram need not be used for the solution of series

circuits. Such a diagram was used in the foregoing explanation merely to facilitate the derivation of Formula (4).

The phase angle $\theta$ is easily computed by trigonometry, once the various voltage values have been ascertained Thus, in Fig. 89,

$$\frac{E_X}{E_R} = \tan\theta, \tag{5}$$

$$\frac{E_X}{E_T} = \sin\theta, \tag{6}$$

$$\frac{E_R}{E_T} = \cos\theta. \tag{7}$$

The trigonometric functions can be computed from whichever known values of voltage are at hand, and the phase angle can be found from a table of natural functions. Because in the explanation above $E_L$ is greater than $E_C$, $E_X$ is an inductive voltage. Therefore, the phase angle $\theta$ is an angle of *lag*, representing the amount by which the current lags the voltage. Had there been a preponderance of capacitive reactance in the circuit, $E_C$ would have been greater than $E_L$, the parallelogram would have appeared in the lower side of the graph, and $E_X$ would have been a capacitive voltage. The angle $\theta$ would then have represented an angle of *lead*.

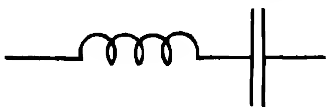


FIG. 90. The theoretical condition of pure inductance and pure capacitance in series.

**Ohm's Law in A-C Circuits.** In a-c circuits, it is customary to regard values of inductance, capacitance, and resistance as being lumped in the circuit. In other words, although inductance, capacitance, and resistance are actually distributed *throughout* a circuit, the mathematical treatment of such circuits is greatly facilitated by assuming that equivalent amounts of these values are concentrated in isolated sections of the circuit. Although such an assumption is not strictly accurate, it is sufficiently precise for most work in radio circuits. Thus, Fig. 90 illustrates an inductance and capacitance in series. If Fig. 90 represents an actual circuit, it is apparent that in addition to the inductance and capacitance shown, there must also be resistance in the circuit. Since by far the larger part of the resistance appears in the wires of which the inductance is wound, it is expedient in theoretical problems to assume that *all* the resistance is in the coil. This value of resistance of the total circuit has the same effect as if it were in series with the circuit. It is therefore represented as a lumped value of resistance in series with lumped values of inductance and capacity. Figure 88 would therefore be a representative equivalent circuit for the practical circuit of Fig. 90.

If there is only resistance in an a-c circuit, Ohm's law can be applied directly in its original form

$$E_R = IR. \tag{8}$$

If an a-c circuit contains inductance, capacitance, and resistance, Ohm's law can be applied to each of the opposition values individually, treating the circuit as though these values were lumped therein. Thus,

$$E_L = IX_L, \tag{9}$$

and

$$E_C = IX_C. \tag{10}$$

Hence, if the current is known, the component voltages for the circuit of Fig. 88 can be computed by substituting in the corresponding form of Ohm's law [Eq. (8), (9), or (10)]. The total voltage is then found by substituting in Eq. (4). In common with all other a-c equations, the values of voltage and current used in a-c Ohm's law equations are effective values.

**Impedance** is the name given to the opposition offered to the flow of alternating current by the combined effects of inductive reactance, capacitive reactance, and resistance. It is expressed in ohms and is represented mathematically by the symbol $Z$.

**Impedance in Series Circuits.** The impedance of a series a-c circuit can be found by substituting in Ohm's law

$$Z = \frac{E}{I}, \tag{11}$$

where $Z$ = impedance in ohms;
$I$ = effective current in amperes;
$E$ = effective voltage obtained from Formula (4) in volts.

The impedance of a series circuit can also be computed directly from the values of reactance and resistance. Since capacitive and inductive reactance are exactly opposite to each other in their effects, they may be represented on the $Y$ axis of Fig. 91 extending in opposite directions. Resistance is represented on the $X$ axis because the effects of inductive and capacitive reactance are each 90° removed from resistance in opposite directions. This vector diagram can be completed to solution in a manner similar to the voltage vector diagram previously discussed. The values of inductive reactance $X_L$ and capacitive reactance $X_C$ are added algebraically to obtain the effective, or net, reactance $X$ (Fig. 91). A parallelogram is constructed having sides $X$ and $R$. Since the reactive vector $X$ and resistance vector $R$ will *always* be at right angles to each other, the parallelogram will *always* have 90° angles. The resultant vector $Z$ is the diagonal of the parallelogram and is the hypotenuse of the right triangle $OAF$. As in the voltage vector analysis, $Z$ can be solved for by application of the Pythagorean theorem. The solution of this vector diagram therefore is summed up by the formula

$$Z = \sqrt{R^2 + (X_L - X_C)^2}. \tag{12}$$

**Values of inductive reactance, capacitive reactance, and resistance can be substituted directly in Formula (12) to obtain the impedance of all series a-c circuits.**

Once the direction of the impedance vector $Z$ has been found, the phase angle $\theta$ can be computed by trigonometry. The phase angle $\theta$ is the same (for a given circuit) as the angle found in the voltage vector

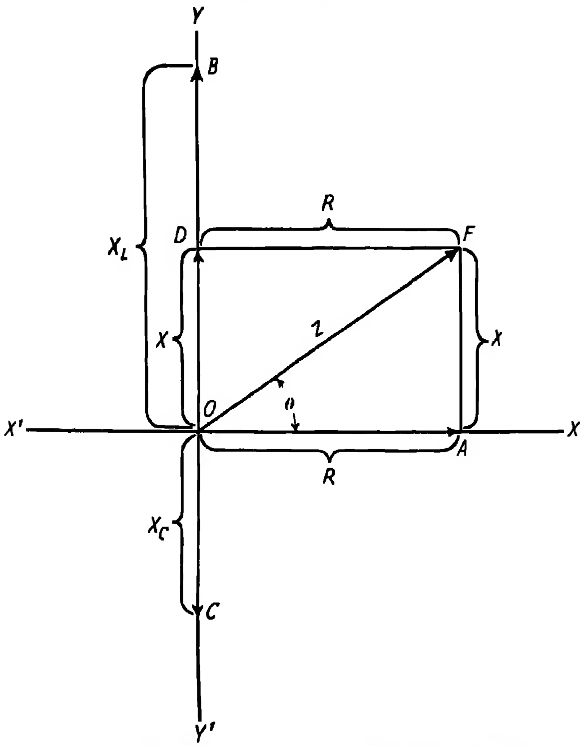



FIG. 91. Impedance vector diagram for series circuits.

diagram, since it is derived as a result of the same forces acting in the same directions. Thus, in Fig. 91, angle $\theta$ can be found by any of the following formulas:

$$\frac{X}{R} = \tan \theta, \tag{13}$$

$$\frac{X}{Z} = \sin \theta, \tag{14}$$

$$\frac{R}{Z} = \cos \theta. \tag{15}$$

When theoretical problems are encountered in which only inductance and resistance are presented in series, the value of $X_C$ in Formula (12) is taken as zero. The remaining values of $X_L$ and $R$ are substituted in the formula, which is then solved for $Z$. Since such a circuit would be preponderantly inductive, the phase angle found by substitution in Formula (13), (14), or (15) would be an angle of *lag*.

Similarly, when a theoretical circuit is to be solved containing only capacitance and resistance, $X_L$ is taken as zero and the remaining substitutions are carried out as above. In such a circuit, of course, the phase angle would be an angle of *lead*. In a circuit containing only inductance and capacitance (a purely hypothetical condition), $R$ is taken as zero. In this case, the disposition of the phase angle would depend upon the relative values of $X_L$ and $X_C$. If $X_L$ exceeds $X_C$, the net $X$ will be inductive, and the current will lag the voltage in the circuit. If $X_C$ exceeds $X_L$, $X$ will be capacitive, and the circuit will have a leading current.

**Problem.** Given a series circuit consisting of a resistance of 4 ohms, an inductive reactance of 4 ohms, and a capacitive reactance of 1 ohm, the applied circuit alternating emf is 50 v. What is the voltage drop across the inductance?

**Solution.** The values of reactance and resistance are substituted in Formula (12) in order to find the impedance, thus,

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{12}$$

Substituting,

$$Z = \sqrt{4^2 + (4 - 1)^2}, \tag{16}$$

$$Z = \sqrt{4^2 + 3^2}, \tag{17}$$

$$Z = \sqrt{16 + 9} = \sqrt{25} = 5 \text{ ohms} \tag{18}$$

According to Ohm's law for a-c circuit

$$I = \frac{E}{Z}. \tag{19}$$

Substituting,

$$I = \frac{50}{5} = 10 \text{ amp.}$$

Then,

$$E_L = IX_L \tag{9}$$

Substituting,

$$E_L = 10 \cdot 4 = 40 \text{ v.} \tag{21}$$

Observe that the voltage across the inductance in the above problem is almost as great as the total voltage impressed across the circuit. This condition should not be confusing, however, since, in series a-c circuits containing inductance and capacitance, component voltages are often encountered that exceed by many times the total impressed voltage.

**Application of this relation is made in practical radio circuits to obtain effective voltage amplifications and is discussed in detail in later chapters on circuit applications.**

## *PARALLEL A-C CIRCUITS*

**Voltage in Parallel Circuits.** In parallel a-c circuits, as in parallel d-c circuits, the voltage is the same across all branches of the circuit. In the circuit shown in Fig. 92, an inductance, a capacitance, and a resistance are connected in parallel to an a-c source. If the voltage across the inductance is represented as $E_L$, the voltage across the capacitance as $E_C$, and that across the resistance as $E_R$, it can be stated mathematically that

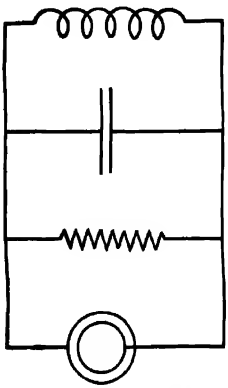


Fig. 92. Inductance, capacitance, and resistance in parallel across an a-c source.

$$E_L = E_C = E_R = E_T, \tag{22}$$

where $E_T$ is the total voltage or the voltage across the generator.

**Current in Parallel Circuits.** Inasmuch as the same voltage exists across all branches of a parallel circuit, it is apparent that the phase displacement caused by the inductance and capacitance of the circuit must be manifested in the branch *currents*. In the resistive branch, current and voltage are in phase that is, current and voltage vectors have the same direction. In the capacitive branch, the current *leads* the voltage by 90° and must be represented as a vector quantity having a direction 90° displaced from the resistive current. In the inductive branch, the current *lags* the voltage by 90° and is therefore represented as a vector quantity having a direction 90° displaced from the resistive current in the direction opposite from the capacitive current. A parallelogram of vectors can be constructed graphically to find the total current. The procedure is identical with that followed in finding the total voltage of a series a-c circuit. The resultant formula for the total current in parallel a-c circuits is

$$I_T = \sqrt{I_R^2 + (I_L - I_C)^2}, \tag{23}$$

where $I_T$ = total circuit current;
$I_R$ = current through resistive branch;
$I_L$ = current through inductive branch;
$I_C$ = current through capacitive branch.

The phase angle in a parallel network is easily computed from vector quantities used to derive Formula (23). Any of the following formulas can be utilized to obtain the phase angle:

$$\frac{I_L - I_C}{I_R} = \tan\theta. \tag{24}$$

$$\frac{I_L - I_C}{I_T} = \sin\theta. \tag{25}$$

$$\frac{I_R}{I_T} = \cos\theta. \tag{26}$$

The displacement of the phase angle (lag or lead) will depend upon the preponderance of inductance or capacity in the circuit. If $I_L$ exceeds $I_C$ the circuit will have lagging current. If $I_C$ exceeds $I_L$ the circuit will have leading current.

**Impedance in Parallel Circuits.** The impedance of a circuit consisting of inductance, capacitance, and resistance in parallel is computed by substitution in the Ohm's law formula

$$Z = \frac{E}{I}, \tag{11}$$

where $Z$ = impedance in ohms,
$E$ = voltage across the circuit in volts;
$I$ = total circuit current in amperes.

According to Formula (22), the voltage $E$ is the same across all branches of the combination. Total current $I$, however, is the vector sum of the individual branch currents and is found by substitution in Formula (23).

It is often desirable to ascertain the impedance of a parallel circuit when neither the voltage nor current is known and only the values of inductive reactance, capacitive reactance, and resistance are given. The procedure in such cases is to *assume* a voltage. The individual branch currents are then computed by the Ohm's law formulas

$$I_R = \frac{E}{R}, \tag{27}$$

$$I_L = \frac{E}{X_L}, \tag{28}$$

$$I_C = \frac{E}{X_C}. \tag{29}$$

The total current is the vector sum of the branch currents and is found by substitution in Formula (23). This resultant current and the assumed voltage are then substituted in Eq. (11) to find $Z$.

Although the values of branch currents obtained by using an assumed voltage have no numerical relation to the currents which may flow in such a circuit under actual operating conditions, the *ratio* of these currents to the voltage across the reactance or resistance is a fixed value for any given circuit and is expressed by Ohm's law. Similarly, the vector

relation between the various branch currents, as expressed by Eq. (23), is a *fixed relation* for any given circuit, regardless of the numerical values of the currents. In other words, for any given circuit with fixed values of $X_L$, $X_C$, and $R$, the value of $Z$ is a constant, regardless of the voltage applied to the circuit and the resultant currents. Therefore, the relationship between the branch currents which flow with a given assumed voltage is the *same* for any value of voltage chosen. It is common practice in working such problems to assume a relatively large value of voltage in order to avoid decimal values and fractions, especially where the reactances involved are large.

**Problem.** A circuit is composed of a 10-ohm resistor, an inductance, and a capacitor connected in parallel. At the frequency at which the circuit is designed to operate, the inductance has a reactance of 4 ohms, and the capacitor has a reactance of 25 ohms. What is the impedance of the circuit?

**Solution.** Assume a voltage of 100 v across the circuit. Then,

$$I_R = \frac{E}{R} \tag{27}$$

Substituting,

$$I_R = \frac{100}{10} = 10 \text{ amp} \tag{30}$$

$$I_L = \frac{E}{X_L} \tag{28}$$

Substituting,

$$I_L = \frac{100}{4} = 25 \text{ amp} \tag{31}$$

$$I_C = \frac{E}{X_C} \tag{29}$$

Substituting,

$$I_C = \frac{100}{25} = 4 \text{ amp} \tag{32}$$

Then,

$$I_T = \sqrt{I_R^2 + (I_L - I_C)^2} \tag{23}$$

Substituting,

$$I_T = \sqrt{10^2 + (25 - 4)^2}, \tag{33}$$

and

$$I_T = \sqrt{10^2 + 21^2} = \sqrt{100 + 441} = \sqrt{541}, \tag{34}$$

$$I_T = 23.2 \text{ amp} \tag{35}$$

According to Ohm's law,

$$Z = \frac{E}{I}, \tag{11}$$

Substituting,

$$Z = \frac{100}{23.2} = 4.3 \text{ ohms} \tag{36}$$

In a theoretical circuit containing only inductance and resistance in parallel, the value of $I_C$ in Formula (23) is taken as zero, that is, the quantity $(I_L - I_C)$ is equal to $I_L$. The remaining values of $I_L$ and $I_R$ are then substituted in the formula, and the solution proceeds as in the above problem. Since such a circuit would be preponderantly inductive, the phase angle found by substitution in Formula (24), (25), or (26) would be an angle of *lag*.

Similarly, when a theoretical circuit is encountered containing only capacitance and resistance, $I_L$ is taken as zero and the remaining substitutions are carried out as above. In such a circuit, the phase angle would be an angle of *lead*. In a circuit containing only inductance and capacitance, $I_R$ is taken as zero. The disposition of the phase angle would depend upon the relative values of $I_L$ and $I_C$. If $I_L$ exceeds $I_C$, the net current will be inductive and will lag the voltage in the circuit. If $I_C$ exceeds $I_L$, the resultant current will be capacitive and will lead the voltage.

## *IMPEDANCE NETWORKS*

**Series-parallel Networks.** It is only in problems concerning hypothetical circuits that values of pure inductance and pure capacitance are encountered. Actually, in practical circuits, a phase angle of exactly 90° is never encountered. Often, however, one particular value of reactance predominates to such an extent that the remaining values (opposing reactance and resistance) are negligible and may be safely disregarded. Strictly speaking, however, *every a-c circuit contains inductance, capacitance, and resistance.*

Consider a coil and capacitor connected in parallel. For rough work, such a combination would be considered as an inductive reactance connected in parallel with a capacitive reactance. Actually, however, the coil possesses shunt capacitance. The capacitance of the coil is due to the electrostatic fields between any one turn of wire and every other turn of wire in it, in other words, any two turns of wire in the coil act as the plates of a capacitor. Since the coil is composed basically of many turns of wire, it possesses resistance due to the ohmic resistance of the wire. A capacitor possesses inductance due to the magnetic field created by the moving electric field of the capacitor. Inasmuch as no dielectric is a perfect insulator, a capacitor also possesses resistance through which leakage currents flow. At high frequencies, the effects of these lesser reactance and resistance values become increasingly important and in many cases cannot be disregarded. The coil and capacitor combination under discussion, therefore, must actually be regarded as a combination of two *impedances* in parallel.

Any complex combination of reactances and resistances can be resolved into a primary network of impedances, either in series or in parallel. Thus,

in Fig. 93(a) an inductance, a resistance, and a capacitance in series are connected in series with a parallel combination of inductance, resistance, and capacitance. Obtaining the total impedance of such a circuit is a question of computing the total impedance of *two impedances* in series. The impedance of the parallel combination of $L$, $C$, and $R$ [$Z_2$ in Fig. 93(a)] is easily computed by the method outlined above. The impedance

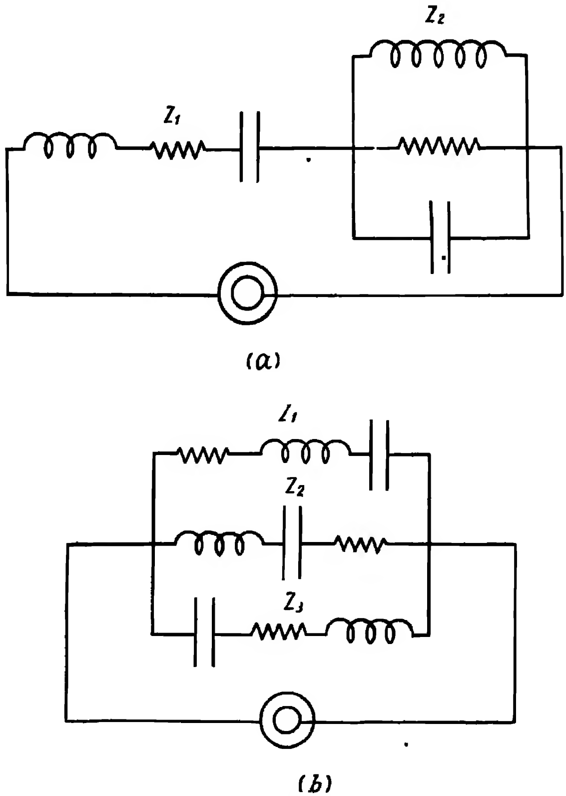



FIG. 93 Complex a c circuits

of the series combination of $L$, $C$, and $R$ [$Z_1$ in Fig. 93(a)] is also easily obtainable. The circuit, therefore, has resolved into two impedances ($Z_1$ and $Z_2$) in series.

Figure 93(b) illustrates another type of complex circuit. Each leg of the combination is a series circuit of inductive reactance, capacitive reactance, and resistance. The impedance of each leg can be found by one of the foregoing methods. The total impedance of the circuit, however, is a problem involving *three impedances* ($Z_1$, $Z_2$, and $Z_3$) in parallel.

Calculating the total impedance of series and parallel impedance networks is a problem involving a more complex notation than that utilized for series and parallel reactance networks. In the reactance problems previously discussed, the forces at work were always 90 or 180° removed from each other, in other words, the parallelogram of vectors (both reactance-vector and current-vector diagrams) was always a *right-angled* figure, permitting the application of the Pythagorean theorem to its solution. Impedances, however, are quantities having phase angles that are *always* some value between 0 and 90°, except in certain cases to be discussed later in the chapter. Thus, the parallelogram of vectors for such circuits will not always have right angles and a different method of solution must be applied.

**Series Impedance Networks.** In a circuit containing several impedances in series, the numerical value of the resultant, or total, impedance may be found by adding the resistive components of all impedances, and the reactive components of all impedances, and then taking the resultant of the two in quadrature. Expressed differently, each individual impedance is resolved into its component resistive and reactive vectors. All the resistive vectors are added separately. All the reactive vectors are added (algebraically) separately. The Pythagorean theorem is then applied to the resultant resistive and reactive components thus obtained to derive the resultant impedance.

Thus, for a circuit containing impedances $Z_1$, $Z_2$, and $Z_3$ in series,

$$Z_1 = \sqrt{R_1^2 + X_1^2}, \tag{37}$$

$$Z_2 = \sqrt{R_2^2 + X_2^2}, \tag{38}$$

$$Z_3 = \sqrt{R_3^2 + X_3^2}, \tag{39}$$

where

$$X_1 = (X_{L1} - X_{C1}); \tag{40}$$

$$X_2 = (X_{L2} - X_{C2}); \tag{41}$$

$$X_3 = (X_{L3} - X_{C3}). \tag{42}$$

It will be noted that the value of $X$ may be either positive or negative, depending in each case upon whether the quantity $(X_L - X_C)$ is preponderantly inductive or capacitive. The value of total impedance $(Z_T)$ of the circuit is given by the expression

$$Z_T = \sqrt{(R_1 + R_2 + R_3)^2 + (X_1 + X_2 + X_3)^2}. \tag{43}$$

Equation (43) is generalized to apply to any number of impedances in series by the formula

$$Z_T = \sqrt{(R_1 + R_2 + R_3 + \cdots)^2 + (X_1 + X_2 + X_3 + \cdots)^2}. \tag{44}$$

Obviously, the values of $R$ and $X$ which compose the individual impedances must be known in order to compute the total impedance. Often

these values are not given in this form. The procedure then is to utilize the trigonometric functions of the phase angle for each impedance to derive the numerical values of the reactive and resistive components. Thus, for the example above, the components are derived as follows:

$$X_1 = Z_1 \sin \theta. \tag{45}$$

$$R_1 = Z_1 \cos \theta. \tag{46}$$

If not given, the phase angle for the total impedance obtained in Formula (44) is derived from the expression

$$\tan \theta = \frac{X_1 + X_2 + X_3 + \cdots}{R_1 + R_2 + R_3 + \cdots}. \tag{47}$$

**Parallel Impedance Networks.** The resultant impedance of several impedances in parallel may be computed in several different ways, all of which may become somewhat involved when a number of impedances are concerned. The reader is urged to review the section on trigonometry in Chap. II before proceeding with parallel impedance networks.

The resultant impedance of several impedances in parallel is computed by substitution in the Ohm's-law formula

$$Z = \frac{E}{I}. \tag{11}$$

The voltage $E$, in common with all other parallel circuits, is the same across all branches of the combination. Total current $I$ is the vector sum of the individual branch currents. Since the individual branch circuits in an impedance are usually not in quadrature with each other, the calculation of total current becomes more complex than in the case of parallel reactance circuits. Perhaps the method is best explained by outlining the procedure for a typical circuit. A series-parallel network is shown in Fig. 94(a). Basically this network consists of two impedances in parallel. The solution of the two series circuits comprising the branch legs has been discussed previously and presents no new problem. Each impedance is represented by the series-circuit formula

$$Z = \sqrt{R^2 + (X_L - X_C)^2}. \tag{12}$$

The total *effective* reactance of such a series circuit is the algebraic sum of the inductive and capacitive reactance and is customarily denoted simply as $X$. Equation (12) can therefore be modified to read

$$Z = \sqrt{R^2 + X^2}, \tag{48}$$

where $X = X_L - X_C$.

If it is assumed in this particular illustration that the upper leg of the parallel circuit is inductive and the lower leg capacitive, a vector diagram

of the currents through the branch legs would appear as in Fig. 94(b). $I_{Z1}$ represents the current in the upper leg and $I_{Z2}$ the current in the lower leg. Since the problem is theoretical, no absolute values are given, and the lengths of the vectors are arbitrary. By completing the parallelogram of forces, the resultant vector $I_T$ (total current through the

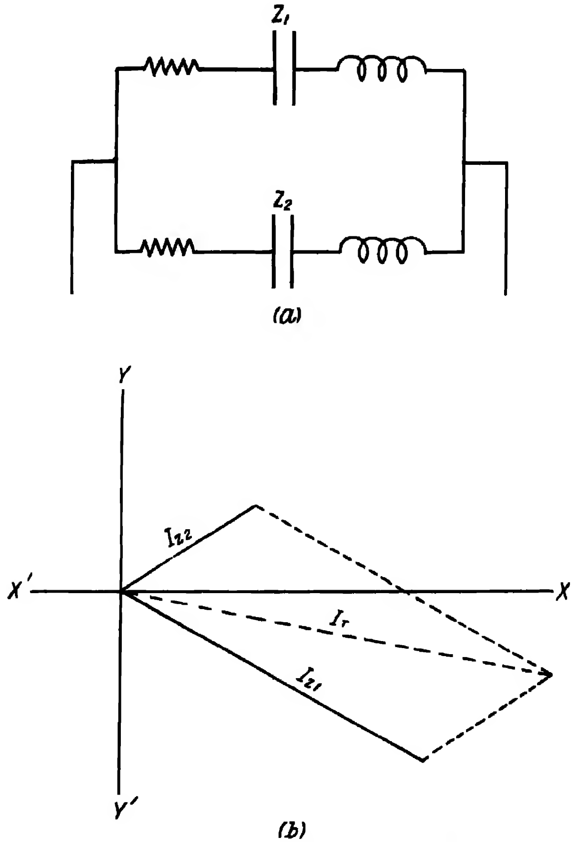



FIG. 94 Current displacement in a parallel impedance circuit

combination) can be found (shown on the diagram by the dashed lines). It is at once apparent that the angles of the parallelogram are not right angles. The mathematical solution of such vectors, therefore, must embrace a different method.

Each branch current, $I_{Z1}$ and $I_{Z2}$, can be assumed to be the resultant of two component currents acting in quadrature. This is illustrated in Fig. 95. $I_{R1}$ and $I_{X1}$ are the components of $I_{Z1}$. Similarly, $I_{R2}$ and $I_{X2}$ are the components of $I_{Z2}$. The component resistive and reactive currents

thus obtained are added algebraically to obtain the *net* effective resistive and reactive currents, so that

$$I_{R1} + I_{R2} = I_{RT}, \tag{49}$$

and

$$I_{X1} + I_{X2} = I_{XT}, \tag{50}$$

where $I_{XT}$ = total effective reactive current,
$I_{RT}$ = total effective resistive current.

It should be noted that since one of the reactive current components ($I_{X1}$) appears *below* the zero line, it takes a negative sign. Since $I_{RT}$

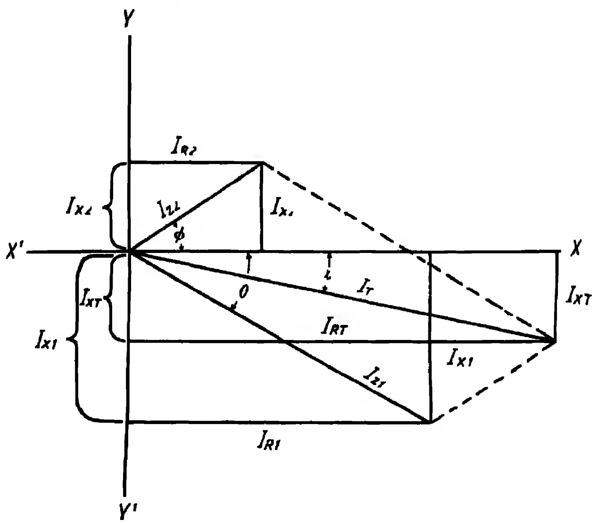



FIG. 95.—Vector currents of Fig. 94(b) resolved into the reactive and resistive components.

and $I_{XT}$ are acting in quadrature, the total current $I_T$ is easily found by application of the Pythagorean theorem

$$I_T = \sqrt{I_{RT}^2 + I_{XT}^2}. \tag{51}$$

Eliminating the graphs, the entire operation can be expressed in the single formula

$$I_T = \sqrt{(I_{R1} + I_{R2})^2 + (I_{X1} + I_{X2})^2}. \tag{52}$$

It should be remembered that the second squared term of Eq. (52) represents the *algebraic* sum of the reactive currents. Where one of the reactive currents is capacitive, as in the example above, this term becomes

$$[I_{X2} + (-I_{X1})]^2 \tag{53}$$

or

$$(I_{X2} - I_{X1})^2. \tag{54}$$

The component current values $I_R$ and $I_X$ are ordinarily not easily obtainable directly and must be computed by trigonometry. Thus, referring to Fig. 95,

$$I_{R1} = I_{Z1}\cos\theta, \tag{55}$$

$$I_{R2} = I_{Z2}\cos\phi, \tag{56}$$

$$I_{X1} = I_{Z1}\sin\theta, \tag{57}$$

$$I_{X2} = I_{Z2}\sin\phi. \tag{58}$$

Substituting these values for their equivalents in Eq. (52), the formula becomes

$$I_T = \sqrt{(I_{Z1}\cos\theta + I_{Z2}\cos\phi)^2 + (I_{Z1}\sin\theta + I_{Z2}\sin\phi)^2}. \tag{59}$$

The value of $I_T$ thus obtained is substituted in the Ohm's law formula of Eq. (11) to obtain the resultant $Z$ of the entire circuit. As in parallel-reactance circuits, if the circuit voltage is not given, one is assumed. The resultant current through each impedance is computed by Ohm's law, utilizing the same voltage throughout the problem.

Since the problem outlined above concerned the case where only two impedances were in parallel, the formula derived [Eq. (59)] holds true for only two impedances as it stands. Nevertheless, the same derivation applies for any number of impedances in parallel. Formula (59) may therefore be generalized for all cases as follows:

$$I_T = \sqrt{\begin{aligned}&(I_{Z1}\cos a + I_{Z2}\cos b + I_{Z3}\cos c + \cdots)^2\\ &+ (I_{Z1}\sin a + I_{Z2}\sin b + I_{Z3}\sin c + \cdots)^2.\end{aligned}} \tag{60}$$

It is apparent that the phase angle of each impedance in a parallel network must be known in order to compute the total, or resultant, impedance. Thus, in the upper leg of the circuit of Fig. 94(a), since it was assumed to be inductive, the current lags the voltage by the angle $\theta$. The value of angle $\theta$ is found as outlined in the section of this chapter devoted to series circuits utilizing Eq. (13), (14), or (15). Impedance $Z_1$ in Fig. 94(a) is then said to be acting "at angle $\theta$." This is written

$$Z_1 \angle \theta \tag{61}$$

and is read, "Impedance $Z_1$ *at* angle $\theta$." Thus, an impedance of 26 ohms acting at a phase angle of 43° is written

$$Z = 26 \text{ ohms} \angle 43° \tag{62}$$

and is read, "an impedance of 26 ohms at angle of 43°." If the circuit is inductive, the phase angle lags; if capacitive, it leads. In the absence of a sign preceding the angle, the angle is taken to be a positive, or lagging, angle. If the angle is leading, it is negative and expression (62) becomes,

$$Z = 26 \text{ ohms} \angle -43°. \tag{63}$$

Often the values of reactance and resistance that make up an impedance are unknown. The impedance phase angle must therefore be given in order to compute the resultant impedance of a number of such impedances in parallel.

The phase angle of the resultant impedance of a parallel network is easily computed from the trigonometric function involving the reactive and resistive components. Thus, referring to Fig. 95,

$$\tan\theta = \frac{I_{XT}}{I_{RT}} \tag{64}$$

In terms of the general formula of Eq. (60), Eq. (64) becomes

$$\tan\theta = \frac{I_{Z1}\sin a + I_{Z2}\sin b + I_{Z3}\sin c + \cdots}{I_{Z1}\cos a + I_{Z2}\cos b + I_{Z3}\cos c + \cdots} \tag{65}$$

**Problem.** Three impedances $10\angle 40°$, $20\angle 29°$, and $25\angle -30°$, are connected in parallel. Find the combined circuit impedance and resultant phase angle.

**Solution.** Assume a voltage of 100 v across the circuit. Then the current through the branch circuits will be

$$I_{Z1} = \frac{100}{10} = 10 \text{ amp} \tag{66}$$

$$I_{Z2} = \frac{100}{20} = 5 \text{ amp} \tag{67}$$

$$I_{Z3} = \frac{100}{25} = 4 \text{ amp} \tag{68}$$

From the trigonometry table in the appendix,

$$\sin 40 = 0.642 \tag{69}$$

$$\cos 40 = 0.766 \tag{70}$$

$$\sin 29 = 0.484, \tag{71}$$

$$\cos 29 = 0.874 \tag{72}$$

$$\sin 30 = 0.500 \tag{73}$$

$$\cos 30 = 0.866 \tag{74}$$

Substituting in Eq. (60),

$$I_T = \sqrt{\begin{array}{l}(10 \times 0.766 + 5 \times 0.874 + 4 \times 0.866)^2 \\ \quad + (10 \times 0.642 + 5 \times 0.484 + (-4 \times 0.500))^2\end{array}} \tag{75}$$

$$I_T = \sqrt{(15.49)^2 + (6.84)^2} \tag{76}$$

$$I_T = \sqrt{239.9501 + 46.7856} \tag{77}$$

$$I_T = \sqrt{286.7357} \tag{78}$$

$$I_T = 16.93 \text{ amp} \tag{79}$$

Substituting in Ohm's law,

$$Z = \frac{E}{I} = \frac{100}{16.93} = 5.9 \text{ ohms.} \tag{80}$$

Then substituting in Formula (65),

$$\tan \theta = \frac{6.84}{15.49} = 0.442, \tag{81}$$

and

$$\theta = 23° \, 51'. \tag{82}$$

The answer therefore is

$$Z = 5.9 \angle 23° \, 51'. \tag{83}$$

## *RESONANCE*

One of the most important phenomena occurring in radio circuits is that of resonance. In both transmitting and receiving networks, resonant circuits are used to build up large voltages or currents at certain desired frequencies while simultaneously discriminating against all other frequencies. The condition of resonance is obtained when the values of inductive and capacitive reactances in a circuit are equal. Inasmuch as these values are 180° out of phase, at resonance they are equal and opposite to each other and therefore cancel.

For any given combination of an inductance and a capacitance there is only *one* frequency to which the circuit is resonant. It was seen in an earlier part of the text that both inductive reactance and capacitive reactance vary with the frequency. Inductive reactance varies *directly* with the frequency, and capacitive reactance varies *inversely* with the frequency. It follows, therefore, that if a given coil and capacitor are connected in a circuit (either series or parallel) and an alternating current of changing frequency is applied to the circuit, there will be but *one frequency* at which the reactance of the coil is equal to the reactance of the capacitor. This frequency is called the **resonant frequency** of the circuit.

Every a-c circuit is resonant at some one frequency. This condition holds whether or not there are lumped values of inductance or capacity in the circuit. At low frequencies the $L$ and $C$ values in the wiring of a circuit are usually so small as to be negligible, that is, the resonant frequency is far removed from any normally used frequency and can be disregarded. At high radio frequencies, however, such distributed values of $L$ and $C$ may become important since the resonant frequencies developed might be in the vicinity of the circuit frequency.

For any given frequency, there is an infinite number of values of $L$ and $C$ which will resonate at that frequency. Thus, if $X_L$ is 10 ohms and $X_C$ is 10 ohms, the circuit is resonant at a certain frequency. If $X_L$

is made 12 ohms by increasing $L$ and $X_C$ is made 12 ohms by decreasing $C$, the circuit is resonant again at the *same* frequency. Obviously, the number of combinations of $L$ and $C$ which cause resonance at the same frequency is infinite.

Since, by definition, the inductive and capacitive reactances of a resonant circuit are equal, it may be stated mathematically that

$$X_L = X_C. \tag{84}$$

Substituting in Eq. (84) the equivalents for these values,

$$2\pi fL = \frac{1}{2\pi fC}. \tag{85}$$

Multiplying both sides of Eq. (85) by $f$,

$$2\pi f^2 L = \frac{1}{2\pi C}. \tag{86}$$

Dividing through by $2\pi L$,

$$f^2 = \frac{1}{(2\pi)^2 LC}. \tag{87}$$

Solving for $f$, we get the formula for the resonant frequency of any series or parallel circuit containing inductance and capacitance:

$$f = \frac{1}{2\pi\sqrt{LC}}, \tag{88}$$

where $f$ = resonant frequency in cycles;
$L$ = inductance in henrys;
$C$ = capacitance in farads;
$\pi$ = the constant 3.14.

**Problem.** What is the resonant frequency of a circuit containing 0.2 h inductance and 0.01 $\mu$f capacitance?

**Solution.** Substituting in Formula (88),

$$f = \frac{1}{6.28\sqrt{0.2 \cdot 10^{-8}}}, \tag{89}$$

$$f = \frac{1}{6.28 \cdot 0.44 \cdot 10^{-4}}, \tag{90}$$

$$f = \frac{10^4}{2.76}. \tag{91}$$

$$f = 3{,}623 \text{ c} \tag{92}$$

**Series Resonance.** In Fig. 96(a) an inductance, a resistance, and a capacitance are connected in series. If this circuit were connected to a source of alternating current, it would be found that at very low frequencies little current would flow, since the reactance of the capacitor is very

high at low frequencies. At very high frequencies, on the other hand, the reactance of the coil is very high, and, therefore, little current would flow. Obviously, maximum current would flow in this circuit at some intermediate frequency somewhere between the very low and very high portions of the frequency spectrum. The exact frequency at which maximum current flow would occur is the *resonant frequency* of the circuit and depends upon the values of $L$ and $C$. At resonance, the inductive and capacitive reactance of the circuit would be equal and opposite and would cancel out. The only opposition offered to the flow of current through the circuit would therefore be the resistance.

Figure 96(b) is an impedance-vector diagram of the circuit of Fig. 96(a) when the circuit is operated at resonance. Since $X_L$ and $X_C$ are equal and opposite, their algebraic sum is zero. The impedance of the circuit therefore, as computed from Formula (12), becomes

$$Z = \sqrt{R^2 - 0}, \tag{93}$$

and

$$Z = R. \tag{94}$$

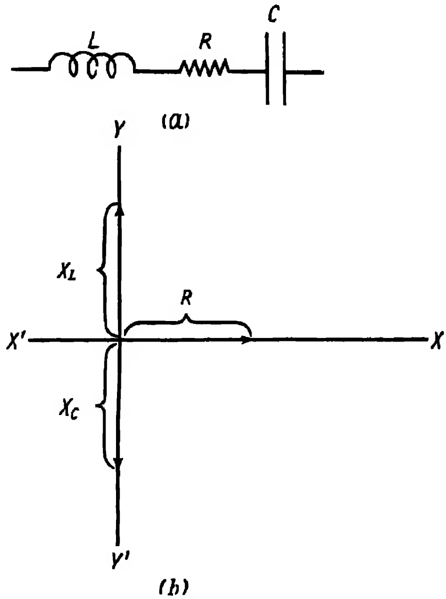



FIG. 96 Series resonance

At any frequency other than the resonant frequency, the values of $X_L$ and $X_C$ will not be equal. Consequently, the algebraic sum of $X_L$ and $X_C$ will not be zero. Equation (12) will then have a reactive component as well as a resistive component, and $Z$ will always be greater than $R$. In series resonance circuits, therefore, it may be said that the impedance is minimum and is equal to the resistance of the circuit at the resonant frequency.

According to Ohm's law, in a series a-c circuit,

$$I = \frac{E}{Z}. \tag{95}$$

$I$ varies inversely as $Z$. As the impedance is decreased, the current is increased. Since at series resonance, the impedance is minimum, it follows that the current is maximum. The current is the same in all parts of a series circuit. Consequently, at resonance, maximum current

flows through the inductive reactance and capacitive reactance of the circuit. By Ohm's law,

$$E_L = IX_L, \tag{96}$$

and

$$E_C = IX_C. \tag{97}$$

The voltage drop across each of the reactances varies directly with the current and directly with the reactance. In the circuits normally used to obtain resonance effects, the current peaks sharply at resonant frequency, and the maximum voltages across $X_L$ and $X_C$ are developed when the maximum current flows through the circuit, that is, at resonance. Actually, it often happens that at resonance the voltage across the inductance or capacitance of a series circuit is greater than the voltage across the entire combination.

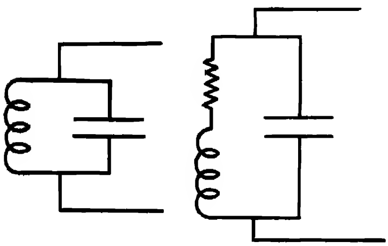

(a) (b)

FIG. 97

Since, in effect, there is only resistance in the circuit at series resonance, the current and voltage will be *in phase*, just as in a simple resistive circuit. The power consumed in such a circuit, therefore, will be the same as for a simple resistive circuit of the same ohmic value and can be found by the formula

$$P = I^2R. \tag{98}$$

For the theoretical condition involving a pure inductance and a pure capacitance in series, $R$ would equal zero at resonance. Consequently, $Z$ would also equal zero (Eq. (94)). The value of $I$ derived from Formula (95) would be infinity.

**Parallel Resonance.** In Fig. 97(a) an inductance and a capacitance are connected in parallel. In an earlier part of this chapter it was seen that the branch currents in such a circuit are 180° out of phase that is, they buck each other. If an a-c source of changing frequency were connected to this parallel circuit, it would be found that at very low frequencies considerable current would flow in the external circuit. The reactance of the capacitor is very high at low frequencies; consequently, the flow of current through the capacitor would be limited. The reactance of the coil at low frequencies, however, is very low, and the current flow through it would be large. Since the flow of current in the inductive leg would be very large compared with the current flow in the capacitive leg, only a small part of the inductive current would be canceled out by the out-of-phase capacitive current. The net current in the external circuit,

therefore, would be large and would be inductive in character, that is, it would *lag* the voltage.

If a v-h-f alternating current were applied to this circuit, it would again be found that the flow of current in the external circuit would be large. At high frequencies, however, it would be found that this current is capacitive in nature, that is, it would *lead* the voltage. The reactance of the condenser would be very low at high frequencies. The current flow through the capacitive leg would therefore be large. The reactance of the coil, on the other hand would be very large at high frequencies, and consequently, very little current would flow through the inductive leg. Only a small part of the capacitive current would be canceled out by the out-of-phase inductive current. The net current flow in the external circuit would therefore be large and capacitive in character.

If the alternating current applied to this parallel circuit is varied from a very low frequency to a very high frequency, it is apparent that at some frequency between these two extremes the inductive and capacitive reactances would be equal. The current in the inductive leg would then be equal and opposite to the current in the capacitive leg, and the result would be zero current in the external circuit. The frequency at which this occurs is the *resonant frequency* of the circuit. When a specific voltage is impressed across a circuit and zero current flows, the impedance of the circuit equals infinity. Thus, according to Ohm's law,

$$Z = \frac{E}{I} = \frac{E}{0} = \infty. \tag{99}$$

It can be stated, therefore that, in a theoretical circuit containing a pure inductance in parallel with a pure capacitance (no resistance in either branch of the circuit) net current flow is zero, and the impedance is infinite at resonance.

In actual practice, of course, all circuits possess some resistance. In a parallel circuit by far the major portion of the total circuit resistance is in the coil. The losses in the average well-designed capacitor are so small that its resistance can usually be neglected. An actual circuit of a coil and a capacitor in parallel would therefore be more accurately represented as a resistance and inductance in series connected in parallel with a capacitor, as shown in Fig. 97(b). Plainly in such a circuit the current will never become zero, for there will always be a flow of current owing to the resistive component in the coil branch of the parallel circuit. At resonance, therefore, the current will pass through a minimum value in a parallel circuit. The impedance, in accordance with the Ohm's-law equation (99), will pass through a maximum value at resonance.

Figure 98 shows the manner in which the impedance $Z$ of a parallel circuit, such as that of Fig. 97(b), varies with the frequency. When such a circuit is placed in series with a constant-current generator, such as a

pentode vacuum tube, the voltage across the parallel combination varies in precisely the same manner as does the impedance.

Figure 99 shows the current vectors for the circuit of Fig. 97(b) when the applied frequency is such that the inductive reactance equals the capacitive reactance—in other words, for the condition of resonance. The capacitive current component $I_C$ is not equal to the inductive current component $I_L$, since $I_L$ does not lag the voltage by a full 90°. For a given circuit, $I_L$ is found by utilizing the laws governing series circuits. Thus,

$$I_L = \frac{E}{Z}, \qquad (95)$$

and

$$Z = \sqrt{R^2 + X_L^2}. \qquad (100)$$

The total current in the inductive leg, therefore, will not be equal to the capacitive current, nor will it be 180° out of phase with the capacitive current. The resultant current flow for the combination, $I_T$, is found by completing the parallelogram of forces, as shown in Fig. 99. It may be noted at once that $I_T$ does not fall on the reference vector line, that is, the current is not in phase with the voltage. It may be shown that the current is not at a minimum for this condition, and, hence, the impedance is not at maximum, although in general it is very nearly maximum. The condition of maximum impedance can be created by varying either $L$, $C$, or the frequency. Similarly, the condition for *unity power factor* (current in phase with the voltage) can be created by varying either $L$, $C$, or the frequency. Since the original definition of resonance for a circuit describes it as the condition where the inductive and capacitive reactances are equal, it follows that at parallel resonance, the impedance is not necessarily at a maximum, nor is the current necessarily in phase with the voltage.

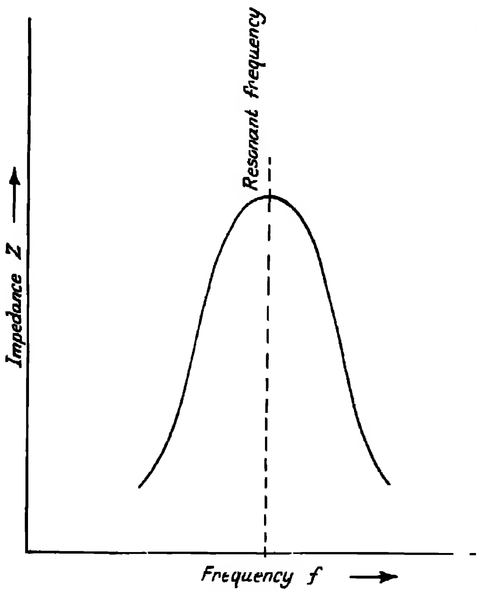


FIG. 98.

This state of affairs has led to several different definitions of parallel resonance which are sometimes confusing to the student. Parallel resonance is sometimes described as the point of maximum impedance, sometimes as the condition at which current and voltage are in phase,

and sometimes as the condition at which inductive reactance equals capacitive reactance. In practical circuits, these three definitions involve resonant frequencies that differ by such a small amount that for all practical purposes the resonant frequency of a parallel circuit may be taken as the frequency at which the inductive and capacitive reactances are equal. Therefore, the frequency that satisfies the relation

$$f_R = \frac{1}{2\pi\sqrt{LC}} \tag{88}$$

will hereafter be taken as the resonant frequency of a parallel circuit in this text. The frequency at which a parallel circuit offers maximum impedance is often called the **antiresonant frequency.** Parallel circuits that are utilized for the purpose of obtaining maximum impedance are called **antiresonant circuits.**

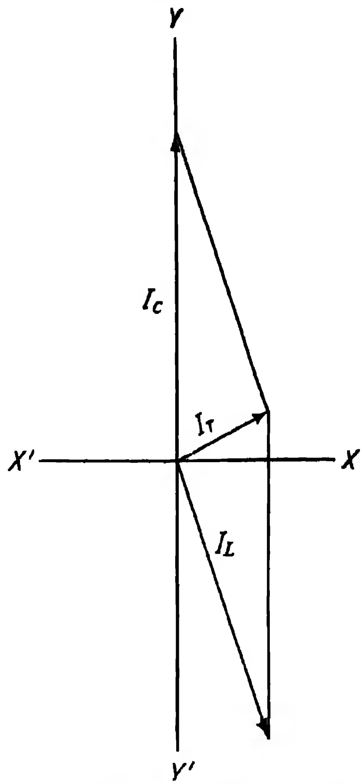



FIG. 99. Vector currents in a resonant parallel circuit having resistance.

## *POWER FACTOR*

The power consumed in watts in a d-c circuit has been shown to be equal to the product of the volts and the amperes. In an a-c circuit, the power consumed can also be shown to be equal to the product of the effective volts and the effective amperes, *provided that the current and voltage are in phase.* This would be the condition for an a-c circuit containing only resistance. If there is any reactance in the circuit in the form of inductance or capacitance, it is apparent from previous treatment of the subject that the current will either lead or lag the voltage, depending upon which value is preponderant. Since the current and voltage in such a circuit do not reach their peak values at the same instant, their product will not be a true representation of the amount of power *consumed* in the circuit. Actually all the power put in the circuit by the generator is *not* consumed by the circuit, and a portion of it is returned to the generator. This condition is due to the back emf of self-induction caused by the collapse of the magnetic field about an inductance and also by the counter current occasioned by the discharge of the capacitance of the circuit. The product

of the volts and amperes of an a-c circuit is called the **apparent power**, or volt-amperes, of the circuit. This would be the value of power obtained by multiplying the readings taken from a voltmeter and an ammeter in a typical a-c circuit. Since the product so obtained is not an accurate indication of the power *consumed* in the circuit, it must be multiplied by a corrective factor called the **power factor** in order to obtain the *true power* of the circuit. The power factor is a function of the phase displacement (the amount by which the current and voltage differ in phase) and is a variable factor that is derived from the phase angle of a given circuit.

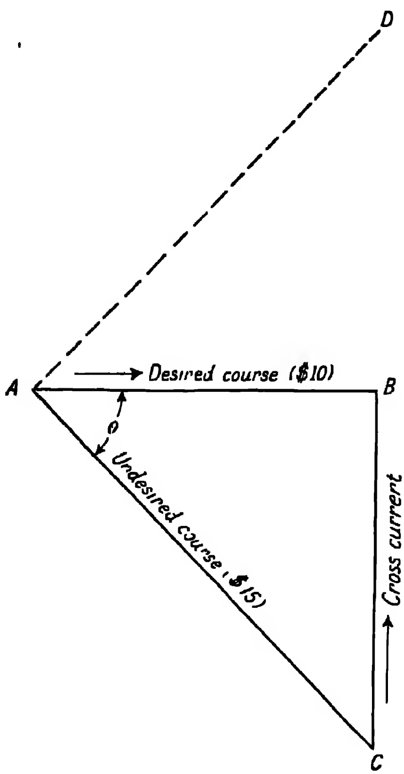



FIG. 100. Analogy of power factor

A conception of power factor is sometimes more easily formed by reference to an analogy. In Fig. 100, a steamboat is proceeding on a course from point $A$ to point $B$. If there are no winds or adverse currents to buck, all the energy expended by the boat as it moves along this course is consumed in overcoming the resistance of the water. Likewise, in an a-c circuit containing only resistance, all the energy supplied to the circuit may be said to be consumed in overcoming the resistance of the circuit. Let us assume that in traversing the distance from $A$ to $B$ under the conditions given above, the steamboat would normally consume $10 worth of fuel.

If, on the other hand, the steamboat were hampered in its movement by a strong crosscurrent at 90° to the desired course (indicated by line $CB$, Fig. 100), it is obvious that the boat would be carried to point $D$ if no steps were taken to overcome this additional type of opposition. The opposition caused by the crosscurrent is analogous to introducing reactance into the a-c circuit previously mentioned. In order to ensure arriving at point $B$, it would be found necessary to head the steamboat in the direction of $AC$. Part of the energy consumed by the boat is expended in overcoming the resistance of the water over path $AB$, and part is expended in overcoming the opposition offered by the crosscurrent.

Obviously, the steamboat expends a greater amount of energy in reaching its ultimate destination (point $B$) under these conditions than it does when no crosscurrent is present. Let us assume that \$15 worth of fuel is expended under these conditions. This \$15 represents the amount of fuel the boat would have consumed had it actually traversed the path $AC$. The undesired path $AC$ may be compared to the power supplied to an a-c circuit containing impedance (combined resistive and reactive opposition).

Under the conditions just outlined, the steamboat, therefore, actually consumed \$15 worth of fuel in moving from $A$ to $B$. Of this amount, only \$10 worth can be said to have been *usefully* consumed, since this would have been the cost had there been no crosscurrent (no reactance). The \$10 may be compared to the *true power* consumed in an a-c circuit; and \$15 may be compared to the *apparent power*. The relation between the two is called the **power factor.** Thus, in the steamboat illustration, the power factor is the ratio between 10 and 15, or,

$$\text{power factor} = \frac{10}{15} = \frac{2}{3} = 0.66 \text{ or } 66\%. \tag{101}$$

For an a-c circuit this equation becomes

$$\text{power factor} = \frac{\text{true power}}{\text{apparent power}} = \frac{P}{EI}, \tag{102}$$

where $P$ = power as read on a wattmeter;
$E$ = voltage as read on a voltmeter;
$I$ = current as read on an ammeter.

In the steamboat illustration this ratio could also be taken as the relation between the undesired path $AC$ and the desired path $AB$, or

$$\text{power factor} = \frac{AB}{AC}. \tag{103}$$

In an a-c circuit, this could be taken as analogous to the ratio between the resistance and impedance, or

$$\text{power factor} = \frac{R}{Z}, \tag{104}$$

where $R$ = resistance of circuit;
$Z$ = impedance of circuit.

By trigonometry,

$$\frac{R}{Z} = \cos\theta. \tag{105}$$

Therefore,

$$\text{power factor} = \cos\theta. \tag{106}$$

By rearrangement of Eq. (102), the following formula for the power consumed in an a-c circuit is obtained:

$$\text{power} = E \cdot I \cdot \text{power factor}. \tag{107}$$

Substituting the equivalent values for power factor taken from Eqs. (104) and (106), Eq. (107) becomes

$$\text{power} = E \cdot I \cdot \frac{R}{Z}, \tag{108}$$

or

$$\text{power} = E \cdot I \cdot \cos\theta. \tag{109}$$

The power factor of any a-c circuit can never be greater than unity, since the cosine of the phase angle can never exceed 1. The condition of unity power factor is obtained in practice when there is only effective resistance in the circuit. This condition occurs in resistance networks, where the capacitance and inductance are so small as to be negligible, and in resonant circuits, where the inductive and capacitive reactances balance out.

**Problem.** Given an a-c circuit having an impedance 10 ohms $\angle 28°$, what is the power factor of the circuit? When an emf of 100 v is impressed across the circuit, what power is consumed in the circuit?

**Solution.**

$$\text{power factor} = \cos 28 = 0.88295 = 88.295\%, \tag{110}$$

and

$$I = \frac{E}{Z} = \frac{100}{10} = 10 \text{ amp} \tag{111}$$

$$\text{power} = 100 \cdot 10 \cdot 0.88295 = 882.95 \text{ w}. \tag{112}$$

When the phase angle of a circuit is positive (circuit preponderantly inductive), the circuit is said to have a **lagging power factor.** When the circuit phase angle is negative, it is said to have a **leading power factor.**

## *CIRCUIT Q*

The selectivity of a circuit, that is, its *sharpness of resonance*, is determined by the ratio of the reactance to the resistance. A figure of merit customarily used to represent this ratio is the letter $Q$. Thus,

$$Q = \frac{2\pi fL}{R}. \tag{113}$$

Since the r-f resistance of a circuit is usually composed almost solely of the coil resistance, the circuit $Q$ is often taken as that of the coil alone. At resonance $2\pi fL = 1/(2\pi fC)$. Therefore, the $Q$ of a resonant circuit may be written as

$$Q = \frac{1}{2\pi fRC}. \tag{114}$$

In a series resonant circuit at resonance, the impedance is equal to the resistance, and the resonant current is

$$I_R = \frac{E}{R}. \qquad (115)$$

where $E$ = the impressed circuit voltage.

The voltage across the inductance is

$$E_L = I_R(2\pi fL).$$

From Eq. (115)

$$E_L = \frac{E}{R}(2\pi fL),$$

$$E_L = E\left(\frac{2\pi fL}{R}\right). \qquad (116)$$

Substituting from Eq. (113),

$$E_L = EQ. \qquad (117)$$

Equation (117) is another way of saying that the resonant rise of voltage in the circuit is equal to $Q$ times the impressed voltage.

It can be shown that when the frequency of the applied voltage in a series circuit differs from the resonant frequency by an amount that is $\frac{1}{2}Q$ of the resonant frequency, the current falls to 0.707 of the maximum, or resonant, value. From this it follows that

$$Q = \frac{f_r}{f_1 - f_2}, \qquad (118)$$

where $f_r$ = resonant frequency;

$f_1$ = frequency higher than resonance at which the current falls to 0.707 maximum,

$f_2$ = frequency lower than resonance at which the current falls to 0.707 maximum.

Similarly, Eq. (118) may be applied to parallel circuits. In this case, the constants become

$f_1$ = frequency higher than resonance at which the *voltage* falls to 0.707 maximum;

$f_2$ = frequency lower than resonance at which the *voltage* falls to 0.707 maximum.

## QUESTIONS AND PROBLEMS*

**1.** What is the current and voltage relationship when inductive reactance predominates in an a-c circuit?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

**2.** Given a series circuit consisting of a resistance of 4 ohms, an inductive reactance of 4 ohms, and a capacitive reactance of 1 ohm. The applied-circuit alternating emf is 50 v. What is the voltage drop across the inductance?

**3.** Should the number of turns of an inductance be increased or decreased in order to raise the resonant frequency?

**4.** Under what conditions will the voltage drop across a parallel-tuned circuit be a maximum?

**5.** A parallel circuit is made up of five branches, three of the branches being pure resistances of 7, 11, and 14 ohms, respectively. The fourth branch has an inductive reactance value of 500 ohms. The fifth branch has a capacitive reactance of 900 ohms. What is the total impedance of this network? If a voltage is impressed across this parallel network, which branch will dissipate the greatest amount of heat?

**6.** If, in a given a-c series circuit, the resistance, inductive reactance, and capacitive reactance are of equal magnitude of 11 ohms and the frequency is reduced to 0.411 of its value at resonance, what is the resulting impedance of the circuit at the new frequency?

**7.** If an alternating current of 5 amp flows in a series circuit composed of 12 ohms resistance, 15 ohms inductive reactance, and 40 ohms capacitive reactance, what is the voltage across the circuit?

**8.** A series circuit contains resistance, inductive reactance, and capacitive reactance. The resistance is 7 ohms, the inductive reactance is 8 ohms, and the capacitive reactance is unknown. What value must this capacitor have in order that the total circuit impedance be 13 ohms?

**9.** If an alternating voltage of 115 v is connected across a parallel circuit made up of a resistance of 30 ohms, an inductive reactance of 17 ohms, and a capacitive reactance of 19 ohms, what is the total circuit current drain from the source?

**10.** What is the meaning of power factor?

## Chapter X

# THE VACUUM TUBE

The radio vacuum tube as it is known today is one of the most extraordinary devices ever conceived by man. It is perhaps the most important single contribution to the world of science since the turn of the century. Without the vacuum tube, the progress of the science of communicating without wires would have been seriously impeded. Radio broadcasting itself owes its inception to the development of this ingenious device. After 40 years of progressive development, there remains much that is not known about the phenomenon of thermionic emission upon which the vacuum tube depends for its operation.

The product of the coordinated efforts of expert engineers and skilled craftsmen, the vacuum tube is a remarkably sensitive and accurate instrument. Of the ninety-odd known elements that comprise the universe, *thirty-seven* are utilized in one form or another in the manufacture of this device. Its future possibilities, even in the light of present day accomplishments, are but vaguely forseen

### *THEORY OF THERMIONIC EMISSION*

**The Edison Effect.** It has been known for over 100 years that when a metal is heated to incandescence, the air in its immediate vicinity becomes a conductor of electricity. This phenomenon came to the notice of Thomas Edison in 1883 during the course of his famous experiments with incandescent lighting. In 1884, Edison discovered that if a conductor were placed in a vacuum with an incandescent metal, the space between that conductor and the incandescent metal became a conductor, and current could be made to flow in the circuit if the two were connected externally (see Fig. 102) It remained for O. W. Richardson to demonstrate that this effect (since called the **Edison effect**) was not due to ionization of air in the vicinity of the conductor. In 1902, Richardson published experimental results confirming his theory that the current flow was due to the actual emission of electrons from the incandescent metal, thus disproving the earlier theory that the air surrounding the metal became conductive. Subsequent experimentation disclosed that the phenomenon was most pronounced in a high vacuum and also that certain metals possessed this unique characteristic to a higher degree than others.

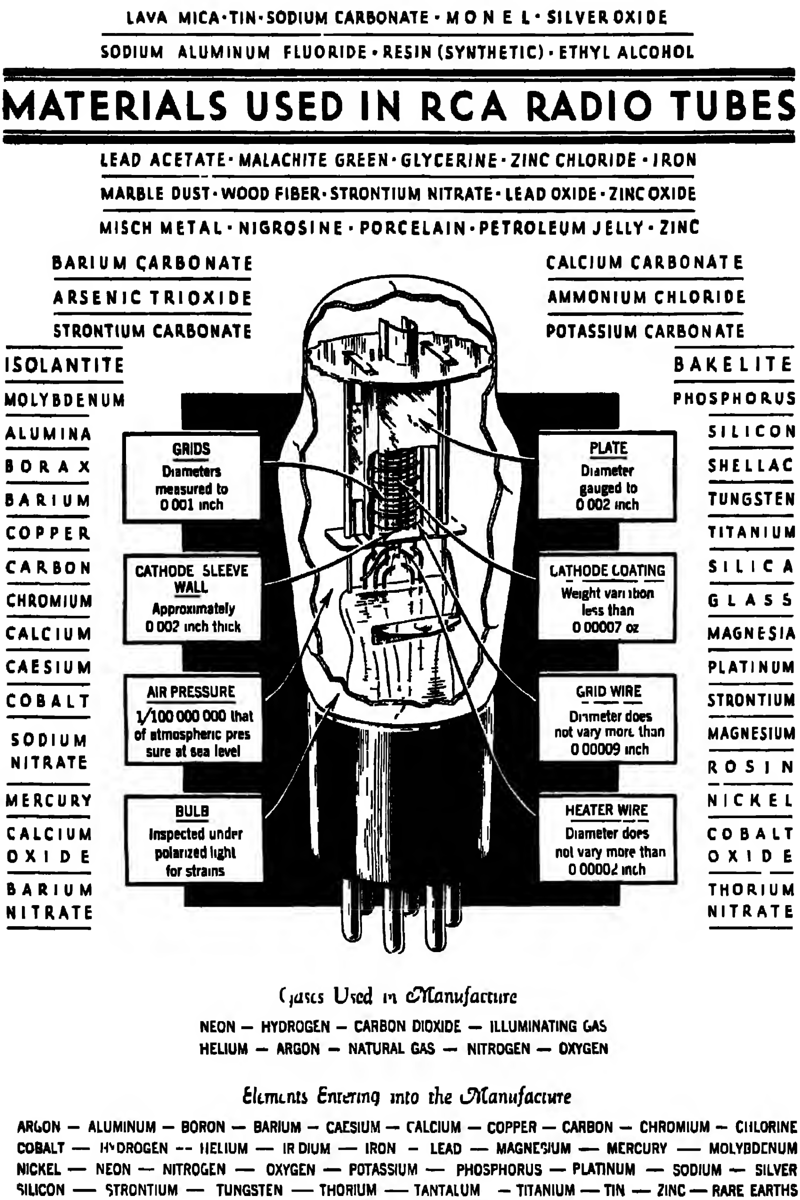

LAVA MICA·TIN·SODIUM CARBONATE·MONEL·SILVER OXIDE
SODIUM ALUMINUM FLUORIDE·RESIN (SYNTHETIC)·ETHYL ALCOHOL
MATERIALS USED IN RCA RADIO TUBES
LEAD ACETATE·MALACHITE GREEN·GLYCERINE·ZINC CHLORIDE·IRON
MARBLE DUST·WOOD FIBER·STRONTIUM NITRATE·LEAD OXIDE·ZINC OXIDE
MISCH METAL·NIGROSINE·PORCELAIN·PETROLEUM JELLY·ZINC
BARIUM CARBONATE
ARSENIC TRIOXIDE
STRONTIUM CARBONATE
CALCIUM CARBONATE
AMMONIUM CHLORIDE
POTASSIUM CARBONATE
ISOLANTITE
MOLYBDENUM
ALUMINA
BORAX
BARIUM
COPPER
CARBON
CHROMIUM
CALCIUM
CAESIUM
COBALT
SODIUM NITRATE
MERCURY
CALCIUM OXIDE
BARIUM NITRATE
BAKELITE
PHOSPHORUS
SILICON
SHELLAC
TUNGSTEN
TITANIUM
SILICA
GLASS
MAGNESIA
PLATINUM
STRONTIUM
MAGNESIUM
ROSIN
NICKEL
COBALT OXIDE
THORIUM NITRATE
GRIDS Diameters measured to 0 001 inch
CATHODE SLEEVE WALL Approximately 0 002 inch thick
AIR PRESSURE 1/100 000 000 that of atmospheric pressure at sea level
BULB Inspected under polarized light for strains
PLATE Diameter gauged to 0 002 inch
CATHODE COATING Weight variation less than 0 00007 oz
GRID WIRE Diameter does not vary more than 0 00009 inch
HEATER WIRE Diameter does not vary more than 0 00002 inch
Gases Used in Manufacture
NEON — HYDROGEN — CARBON DIOXIDE — ILLUMINATING GAS
HELIUM — ARGON — NATURAL GAS — NITROGEN — OXYGEN
Elements Entering into the Manufacture
ARGON — ALUMINUM — BORON — BARIUM — CAESIUM — CALCIUM — COPPER — CARBON — CHROMIUM — CHLORINE
COBALT — HYDROGEN — HELIUM — IRIDIUM — IRON — LEAD — MAGNESIUM — MERCURY — MOLYBDENUM
NICKEL — NEON — NITROGEN — OXYGEN — POTASSIUM — PHOSPHORUS — PLATINUM — SODIUM — SILVER
SILICON — STRONTIUM — TUNGSTEN — THORIUM — TANTALUM — TITANIUM — TIN — ZINC — RARE EARTHS

FIG. 101

The great work of such scientists as Richardson, Langmuir, Wilson, Thomson, and many others in this field has provided the foundation for the study of electronics. It can now be stated authoritatively that certain metals, when heated to incandescence in a vacuum, emit, or evaporate, electrons Since electrons are *negative* particles of electricity, they are attracted to any body having a positive charge. Thus, when a conductor is introduced in the same vacuum as an incandescent emitter, the electrons from the latter are strongly attracted to the conductor, constituting a flow of current between the two elements. Since the cold secondary conductor must be positive with respect to the source of electrons, it is called the **anode.** It is also commonly referred to as the **plate,** since in its early form it was a rectangular plate. The emitter of electrons, that is, the metal that is heated to incandescence, is called the **cathode,** since it is negative with respect to the plate or anode. Modern vacuum-tube cathodes are usually constructed of thoriated tungsten or oxide-coated Konel metal. The thoriated cathode is composed of tungsten in which about ½ per cent of thorium oxide and a small amount of carbon have been dissolved. The entire cathode is then coated with an extremely thin layer of the metal thorium. The Konel-metal cathode utilizes a base of Konel metal upon which successive coatings of barium and strontium oxides have been applied. Konel metal is an alloy of nickel, cobalt, iron, and titanium.

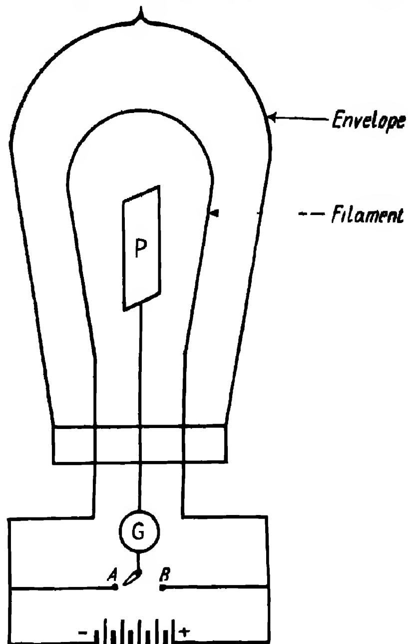


FIG 102 The Edison effect. When the switch is connected to point *A*, the galvanometer does not register; when connected to point *B*, a flow of current is indicated by the galvanometer

**The Fleming Valve.** One of the first to find practical application for the Edison effect in thermionic tubes was the British scientist Fleming. The conductivity of a vacuum tube having a cold anode and a hot cathode is unilateral, since the electrons from the cathode are attracted to the anode only when the latter is positively charged. When the anode is made *negative* with respect to the cathode, the electrons are repelled from it, and zero current flows in the circuit. Fleming utilized this "rectifying" characteristic of the vacuum tube for the detection of

damped h-f waves. The famous Fleming valve patent was granted to him for this application in 1905. The term "valve" is of British origin and in this sense is synonymous with the American word "tube." The following discussion will bring out the appropriateness of the term.

The principle of thermionic rectification as first applied by Fleming is widely utilized in modern radio equipment for the purpose of converting 60-c alternating current to direct current. As will be seen in the forthcoming treatment of the thermionic amplifier, direct currents are necessary for the successful functioning of a vacuum tube as an amplifier.

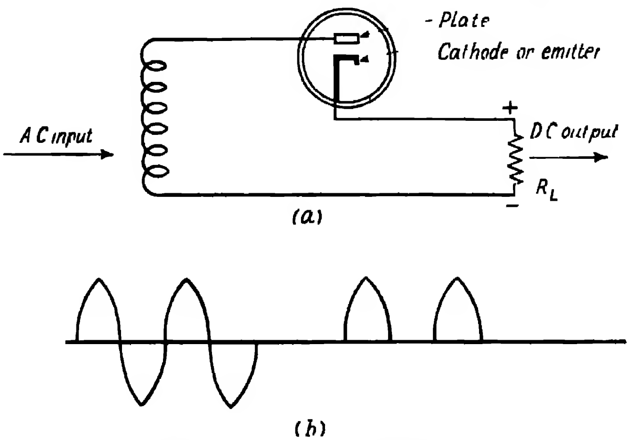



FIG. 103. (a) Elementary rectifier circuit (b) Wave forms for the circuit of (a)

The circuit of a typical 60-c vacuum-tube rectifier is shown in Fig. 103, with a graph of the input voltage and output current. The cathode heating circuit has been omitted since it takes no active part in the rectifying action. Analysis of the input-voltage form discloses that for the first alternation the plate is positive with respect to the cathode. During the first 90° of the cycle, the charge on the plate varies from zero to maximum (peak voltage). As this positive charge on the plate increases, the number of cathode electrons that are drawn to the plate increases in proportion, owing to the progressively greater intensity of the positive charge. The current that flows in this *plate circuit*, therefore, varies in accordance with the impressed sine wave of voltage. During the following 90° period, the impressed voltage falls from maximum to zero, and, consequently, the plate-circuit current falls to zero. On the following alternation, the charge applied to the plate is opposite in polarity. The electrons emitted from the cathode are repelled by the negatively charged plate, and none of them arrives at the plate. Hence, no plate current flows. Since the impressed voltage is negative for the entire alternation, the plate current is zero throughout this entire alternation. On the

following cycle, the identical process is repeated resulting in an output current form as graphically depicted in Fig. 103(b).

It will be noted that this output current consists of periodic pulses of direct current. This pulsating direct current is filtered (by means described in a later chapter) until a smooth nonpulsating direct current is obtained for practical use. The losses during the entire process are considerable, and the amplitude of the resulting smooth output current is appreciably less than the peak output value of the rectifier.

**The De Forest Audion.** The next major improvement in the development of the vacuum tube was made by Dr. Lee De Forest. De Forest inserted a third element, which he called a **grid,** between the cathode and plate of the vacuum tube. A means was thus provided for controlling the flow of electrons in the **audion,** as he called the device. This control grid enables the vacuum tube to function as both an amplifier and an oscillator. The tremendous potentialities of the three electrode tube, or triode, and the rapid development of the science of radio that resulted from its invention earned for De Forest the sobriquet of "the father of radio." De Forest was granted a patent for his triode in 1907.

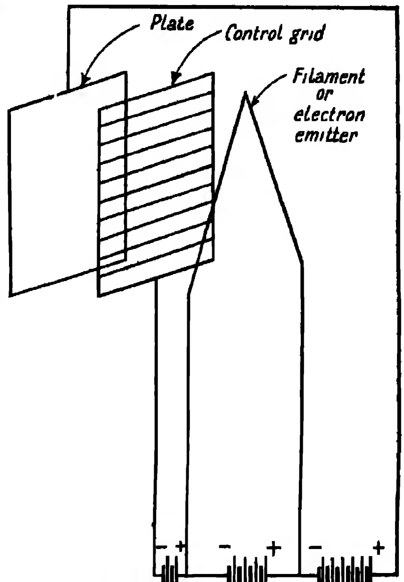


FIG. 104. Elementary triode tube circuit.

The control grid received its name from its resemblance to a gridiron (see Fig. 104). It is composed of several strands of very fine wire in the shape of a gridiron and is inserted in the electron path between the emitter and the plate. If a steady positive charge is applied to the plate, there will be a continuous flow of electrons from the filament, or emitter, to the plate and a resultant flow of plate current in the external circuit which is essentially constant in amplitude. Since the wires of the grid are comparatively widely separated, the electrons encounter little opposition from the physical structure of the grid.

If the control grid is made positive with respect to the emitter, the velocity of the electrons that leave the emitter is greatly increased and a greater number of them are attracted in the direction of the grid and plate. Since the grid is much closer to the cathode than the plate, a positive charge on the grid has a correspondingly greater effect upon the

electrons. As the electrons pass through the grid, they come under the influence of the more powerful plate charge, and most of them continue on to the plate as before. A small portion of the electron flow, however, is attracted to the grid, causing a flow of current in the external grid-cathode circuit (see Fig. 104). Nevertheless, the net result of making the grid positive is an increase in plate current.

This control-grid action is perhaps more easily followed if one considers the action of the emitter at first with no external charges in the vicinity, that is, no charge on plate or grid. With no positive charge about, the electrons emitted from the filament hover in the vicinity of the filament. As the emission continues, the electrons that were emitted earlier have a tendency to repel those emitted later, since the outer electrons constitute a negative charge in themselves. The result is a random cloud of electrons in the vicinity of the filament that forces some electrons back to the filament. This is called the **space charge,** since it constitutes a negative charge in space. When a positive potential is applied to the plate, a portion of this positive charge is utilized in neutralizing the space charge. The remaining positive charge can be thought of as active in accelerating the flow of electrons as they are attracted to the plate. The neutralization of the space charge allows electrons to be emitted from the filament practically unimpeded by any space charge.

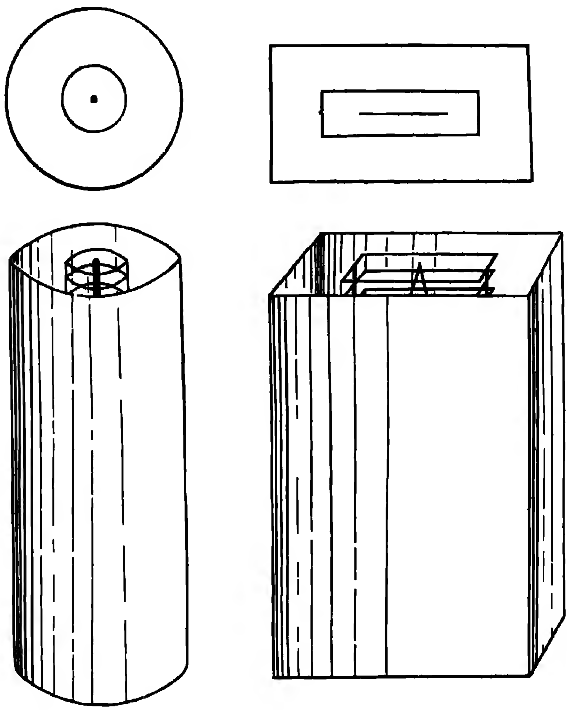
Since the grid is physically nearer the filament than is the plate, a small positive potential applied to the grid will have the same effect as a large positive potential on the plate. The grid therefore takes over the job of neutralizing the space charge. The positive potential of the grid so near the source of emission speeds up the escape of the electrons and prevents the building up of a high negative space charge. The positive plate charge is consequently free to carry on its primary function of attracting the electrons.

If the grid is made *negative* with respect to the filament, the effect is the same as if an increase had been made in the negative space charge. The grid then acts as a barrier that resists the passage of electrons. Some electrons do succumb to the attraction of the positive plate charge, but the net result is a decrease in plate current. Since the grid is negative, no electrons at all are attracted to it,* and no current flows in the external grid-cathode circuit. If the negative charge on the grid is made sufficiently high, the equivalent space charge becomes so great that no electrons reach the plate and plate current becomes zero. When this condition is reached, the tube is said to have been "blocked."

It was not long before the original style of tube structure was improved upon. The inefficiency resulting from the use of planar elements as shown in Fig. 104 disclosed that a large portion of emitted electrons hovered

* A few high-velocity electrons *are* emitted by the heated cathode. These high-velocity electrons strike the grid and are returned to the cathode even at high values of negative bias.

about the filament on the side away from the plate. They were thus not subject to the positive charge of the plate to so great a degree, and many of them never reached the plate at all. Modern tube design makes use of grid and plate elements that entirely surround the emitter. Such elements are often cylindrical in form, as in Fig. 105(a), and frequently are



(a) (b)

FIG. 105 Modern triode construction (a) Cylindrical (b) Rectangular.

manufactured in the rectangular shape shown in Fig 105(b). This type of construction utilizes to the fullest extent the emission of the filament.

The batteries originally employed to operate triode vacuum tubes were given separate designations to distinguish between their functions. The battery used to heat the filament, or emitter, to incandescence was called the "A battery." The battery supplying the plate with its necessary positive charge was called the "B battery." As will be seen later, a third was sometimes employed to give the grid a specific average potential. Such a battery was called a "C battery."

The use of batteries was soon outmoded when the thermionic rectifier was developed to the point where 60-c alternating current could be

converted to direct current economically and practically. The B and C batteries were replaced by power supplies, often called **power packs.** They retained their original designations, however, and power packs are still called B packs and C packs, depending upon their application.

The replacement of the A battery presented certain complications, however. Because of the comparatively large current drain of the filament of a vacuum tube, it was not practical to construct a rectifier to supply this current, especially when, as was almost always the case, there were several vacuum tubes to be supplied. Alternating current, it was found, could not be applied directly to a filament for heating purposes because the variations in amplitude of the alternating current caused a proportional variation in the electron emission. The resulting variation in plate current occurred at double the frequency of the alternating current usually 60 c, causing a hum of 120 c. The hum level was so high that it precluded the use of this type of A supply.

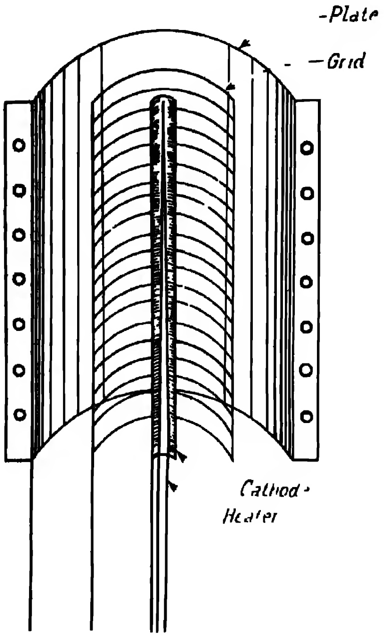



FIG. Cutaway view of triode tube showing indirect heater construction

In later years the indirectly heated type of emitter was developed. Here the construction employs a hot filament, surrounding which the actual emitter is built. The emitter is electrically insulated from the **heater,** as this type of filament is called. By conduction and radiation the heater heats the emitter, called the "cathode," to incandescence, and the latter thereupon emits electrons. Alternating currents may be applied directly to the heater with such an arrangement without derogatory 60 c hum components in the plate current. The variations of the amplitude of the alternating current are so rapid, compared with the time lag in heat transmission, that substantially no variation in electrons emitted from the cathode occurs. Figure 106 is a cutaway illustration of a typical indirectly heated type of triode.

## *THE VACUUM TUBE AS AN AMPLIFIER*

**General Theory of Amplification.** The action of a triode as an amplifier is best understood by analyzing the action of such a tube as the applied

voltage is varied. Figure 107 illustrates an elementary triode amplifier circuit. The heater circuit is omitted, since it takes no part in the amplifier action. The grid-cathode circuit is called the **input circuit** because normally the signal voltage to be amplified is impressed across this circuit. The plate-cathode circuit is called the **output circuit** because the amplified output voltage is taken from this circuit. The output load is represented by the resistor $R_L$. The C battery is so arranged in the input circuit that the grid can be made either positive or negative with respect to the cathode at will, thus simulating any condition of signal voltage.

If the arm of the potentiometer $P$ in Fig. 107 is adjusted to the mid-point of its range, there will be zero impressed voltage across the input circuit that is, the grid will be neither positive nor negative with respect to the cathode. The plate current as indicated on the ammeter $A$ will

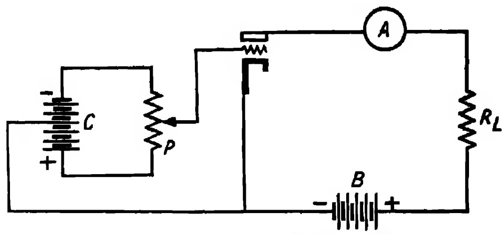



FIG. 107 Experimental vacuum tube circuit

remain at a constant value. Assume this value to be 10 ma to facilitate explanation. If the voltage of the B battery is increased by 10 v, it follows that the plate current will be increased. Let us assume that the amount of plate-current increase for this condition will be 1 ma, then 11 ma of current will flow in the plate circuit. Conversely, decreasing the plate voltage a similar amount from normal will cause the plate current to decrease by the same amount (1 ma)

If the plate voltage is returned to normal and the grid voltage is varied, it will also be found to have an effect on the plate current. It will be found that making the grid negative with respect to the cathode by only 1 v will decrease the plate current by 1 ma Conversely, making the grid positive by 1 v will increase the plate current by 1 ma Varying the grid voltage only 1 v in this case has the same effect upon the plate current as varying the plate voltage 10 v. Thus, if an alternating voltage with a peak value of 1 v were impressed across the input circuit of this tube, the effect upon the plate current would be the same as if an alternating voltage with a peak value of 10 v were impressed across the output circuit. Expressed in another way, the impressed 1 v of alternating voltage has been "amplified" by the tube 10 times. It should be remembered that the numerical values used above have been arbitrarily chosen and do not refer to any specific type of tube.

Actually, the term "amplifier" is a misnomer. The impressed signal voltage, instead of being amplified, was *reproduced* in the plate circuit with much greater amplitude. The comparatively powerful current of the B battery was given the identical characteristics of the impressed grid voltage. The vacuum tube really acts as a valve that controls the amplitude of the plate current in accordance with the amplitude of signal voltage on its grid, and from this analysis the British use of the word "valve" to mean vacuum tube received its derivation. In commercial telephone practice, vacuum-tube amplifiers are often called "repeaters." In this respect both valve and repeater are more accurate representations of the action of a vacuum tube.

**Saturation.** With a given fixed heater voltage (fixed amount of electron emission from the cathode), more and more electrons will be attracted to the plate as the plate voltage is increased. As the increase is continued, however, a point is eventually reached where *all* the electrons being emitted from the cathode are being attracted to the plate. Beyond this point further increase in plate voltage does not increase the plate current. This point is called the **plate saturation point** of the tube.

If the plate voltage is fixed at a certain operative value and the heater voltage is increased, another type of saturation occurs. At zero heater voltage, no electrons are being emitted from the cathode and plate current is zero. As the heater voltage is increased, the plate current rises, owing to the emitted electrons that are attracted to the plate. Soon, however, the cathode is emitting electrons faster than the plate can attract them. The surplus electrons (those which do not get to the plate) form a negative space charge which further hampers the flow of electrons to the plate. Finally, a point is reached where further increasing the heater voltage does not increase the plate current, because of the amount of space charge which has been built up. This point is called the **cathode saturation point** of the tube.

After the cathode saturation point has been reached, the only way to increase plate current is to increase the plate voltage, thus neutralizing part of the space charge. When neither increasing the heater voltage nor increasing the plate voltage of a tube will increase the plate current, the tube is said to be operating at **peak saturation point.** Peak saturation is therefore due to the combined effects of cathode and plate saturation.

## *VACUUM-TUBE CHARACTERISTICS*

**Amplification Factor.** The amplification propensities of a given tube depend primarily upon the tube-element structure. The proximity of the grid structure of a tube to the cathode governs the extent of control exercised by a grid charge over the electron flow. Similarly, the proximity of the plate to the cathode governs the extent of control that a plate

charge has upon the electron flow. Hence, it follows that the interelectrode dimensions of a tube (plate to cathode and grid to cathode, as well as grid wire spacing) regulate the relative degree of control exercised by the two elements (plate and grid). Thus, if the plate is twice as far removed from the cathode as is the grid, it is obvious that a given charge on the grid will have a greater effect upon the electron flow (and resultant plate current) than will the same charge on the plate. In the illustration cited earlier in this chapter it was found that a 1-v charge on the grid had the same effect on the plate current as did a 10-v charge on the plate. Stated in another way, it could be said that a 1-v charge on the plate of this tube would have only *one tenth* as much effect on the plate current as a 1-v charge on the grid. Inasmuch as the grid (input) voltage has the same effect on the plate (output) current as does a voltage 10 times greater, it is said that the tube has a gain of "10 times." The expression for the amplification that it is theoretically possible to realize from a tube is called the **amplification factor.** Thus, the tube mentioned above has an amplification factor of 10. Amplification factor may be defined as the *change in plate voltage necessary to produce a change in plate current divided by the change in grid voltage necessary to produce the same change in plate current.* Expressed mathematically,

$$\text{amplification factor} = \frac{dE_p}{dE_g} \text{ for the same } dI_p, \tag{1}$$

where $dE_p$ = change in plate voltage;
$dE_g$ = change in grid voltage,
$dI_p$ = change in plate current.

(The letter $d$ is the customary mathematical symbol denoting increment or change.)

The mathematical symbol for amplification factor is the Greek letter $\mu$, pronounced "mu" as in *music*.

Amplification factors of over 1,000 are not unusual in modern tubes. As will be seen later, however, such high amplifications are not obtained in actual practice, because of the limiting circuit factors that must be taken into consideration.

**Plate Resistance.** The plate resistance is an important tube constant which must be considered in vacuum-tube circuit design, regardless of the tube application. It should not be confused with the d-c resistance of a tube. The latter is obtained from the Ohm's law relation between the d-c voltage and current in the plate circuit. It is well to stop at this point and consider the nature of the complex plate current in a tube circuit when a signal voltage is impressed on the grid.

With no signal voltage present in the input circuit, the plate current of a vacuum tube would be uniform in amplitude, as is illustrated by the graph in Fig. 108(a). If a pure sine wave of alternating voltage were

impressed on the input circuit of the tube, the plate current (assuming no distortion in the tube) would appear as in the graph of Fig. 108(b). The variations in plate-current amplitude would be identical in form to the original impressed sine wave. Despite the fact that this current is unidirectional (it is still direct current), the variations in amplitude will have the same effect on any a-c-operated device as would a pure alternating current of the same amplitude variation. Thus, if the pulsating direct current of Fig. 108(b) were passed through the primary of a transformer, the output of the secondary side of the transformer would be alternating voltage, since the transformer depends for its operation upon the *variation* in amplitude of current, *not* upon the reversal of polarity.

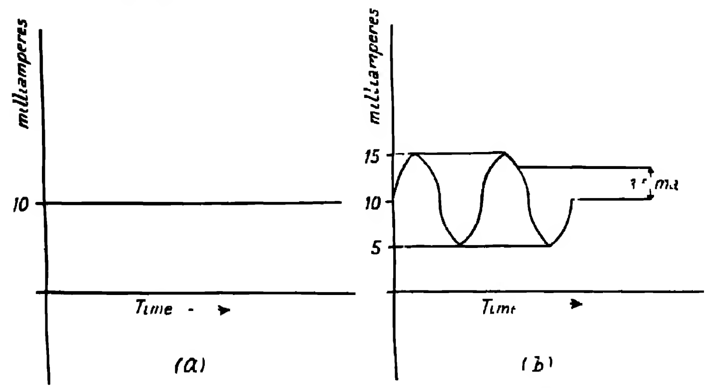



FIG. 108. Nature of the current in a vacuum-tube plate circuit. (a) Plate current with no signal voltage (b) Plate current with sine-wave signal voltage.

The value of alternating voltage obtained from the secondary would be the same as that obtained if an *alternating* current of the same effective value as the *pulsating portion* of the direct current had been applied to the primary.

It should be remembered that the smooth portion of the plate current takes no active part in the output operation. Thus, if the normal smooth plate current (no grid voltage present) were 10 ma and an exciting grid voltage caused maximum and minimum currents of 15 ma and 5 ma alternately (see *Fig.* 108[b]), the *effective* varying portion of the plate current would be 3.5 ma. This value is obtained by multiplying the peak variation *from normal* (5 ma) by the factor 0.707, and this portion of the plate current is called the "a-c component" of the plate current because it is capable of accomplishing the same amount of work as an alternating current of the same effective value. Thus, in the example above, the a-c component of the plate current will accomplish the same amount of work as an alternating current having an effective value of 3.5 ma. The *d-c component* of plate current refers to the normal direct current flowing in the circuit (nonpulsating).

The *d-c resistance* of a tube (the resistance between plate and cathode) is obtained by simply applying Ohm's law with no other resistance in the circuit. Thus,

$$R_{dc} = \frac{E_p}{I_p}, \tag{2}$$

where $E_p$ = d-c plate voltage applied to the tube;
$I_p$ = d-c plate current with no grid voltage present.

This d c resistance is by no means a constant, for it depends upon the particular value of grid voltage as well as the plate current, and has very little significance in itself in practical radio problems.

The *dynamic*, or *a-c*, *plate resistance*, on the other hand, is a very important quantity and is approximately constant for a given operating range of a tube. It depends on the *change* in plate current resulting from a *change* in plate voltage. Mathematically, this dynamic plate resistance, commonly called simply **plate resistance,** may be expressed:

$$R_p = \frac{dE_p}{dI_p}, \tag{3}$$

where $R_p$ = plate resistance of the tube;
$dE_p$ = change in plate voltage;
$dI_p$ = change in plate current resulting from $dE_p$.

For this definition, the changes in plate voltage and plate current are to be slow changes. If the changes are rapid, as is the case when the tube is operating at radio frequencies, the plate and cathode act as a capacitor introducing a shunt capacitance in the circuit. In this case the opposition to a-c plate current is an impedance, rather than a pure resistance, and is called **plate impedance.**

**Mutual Conductance.** It would seem at first glance that the amplification factor of a tube is the ideal constant to employ when comparing the relative efficiencies of several tubes. It is, of course, desirable to have the amplification factor as high as possible when choosing a tube as an amplifier Nevertheless, the maximum amount of amplification (as evidenced in the a-c plate current component) is not necessarily realized from the tube having the highest amplification factor. The plate current (a-c component) is a function not only of circuit voltage but also of the resistance.

An equivalent circuit for the plate circuit of a vacuum tube is shown in Fig. 109(a). The a-c voltage impressed across this circuit is equal to the impressed grid voltage $E_g$ multiplied by the amplification factor of the tube, $\mu E_g$. The a-c component of the current that flows in this plate circuit is, then, according to Ohm's law, equal to

$$I_{ac} = \frac{\mu E_g}{R_p + R_L}, \tag{4}$$

where $R_p$ = plate resistance of the tube;
$R_L$ = resistance of the load to which the tube is coupled.

It is apparent that the amplitude of the a-c plate-current component depends not only upon the amplification factor of the tube but also upon the plate resistance. Thus, if two tubes have amplification factors of 250 and 350, respectively, the latter tube will not necessarily produce a larger output for a given input voltage. If the latter tube has a very large plate resistance, its output might conceivably be even less than the former. Obviously, the amplification factor is not sufficiently indicative of a tube's efficiency to serve as a "figure of merit" for the tube.

To fulfill the need for such a figure of merit, the term **mutual conductance** was created. It was conceived by Hazeltine in 1918 and is often called **transconductance.** Mutual conductance combines amplification

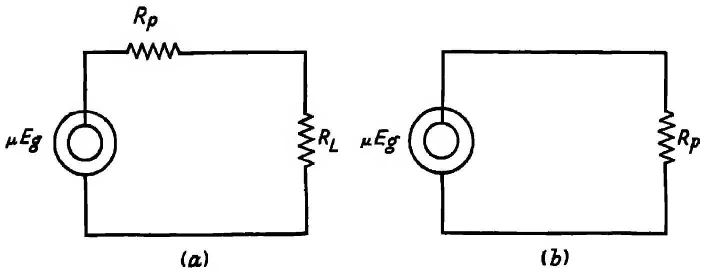



FIG. 109. Vacuum-tube equivalent circuits

factor and plate resistance in a ratio which is representative of a tube's circuit efficiency. Mathematically, it is equal to

$$G_m = \frac{\mu}{R_p}, \tag{5}$$

where $G_m$ = standard symbol for mutual conductance in mhos:
$R_p$ plate resistance of the tube;
$\mu$ amplification factor of the tube.

The following conclusions may be drawn: the higher the mutual conductance of a tube, the higher its efficiency, the lower the plate resistance of a tube for a given amplification factor, the higher the mutual conductance, the higher the amplification factor for a given plate resistance, the higher the mutual conductance.

The mutual conductance of a tube may be measured directly with the set-up in Fig. 107 if the plate load $R_L$ is equal to zero. The equivalent circuit for this condition is illustrated in Fig. 109(b). For this circuit, Eq. (4) becomes

$$I_{ac} = \frac{\mu E_g}{R_p} = G_m E_g, \tag{6}$$

where $E_g$ is the a-c grid voltage. $G_m$ is then the ratio of the a-c plate current to the a-c grid voltage when $R_L$ equals zero.

The mutual conductance may also be expressed in terms of current and voltage changes. According to Eq. (1),

$$\mu = \frac{dE_p}{dE_g} \text{ for constant plate current.} \tag{1}$$

According to Eq. (3),

$$R_p = \frac{dE_p}{dI_p} \text{ for constant grid voltage.} \tag{3}$$

Substituting the equivalent values for $\mu$ and $R_p$ obtained from Eqs. (1) and (3) in Eq. (5),

$$G_m = \frac{\dfrac{dE_p}{dE_g}}{\dfrac{dE_p}{dI_p}}, \tag{7}$$

and

$$G_m = \frac{dI_p}{dE_g} \text{ for constant plate voltage.} \tag{8}$$

Mutual conductance is seen, therefore, to be a direct ratio of the plate-current change for zero plate load to the grid voltage change. Mutual conductance is usually expressed in micromhos.

**Characteristic Curves.** The action of the vacuum tube is itself a phenomenon. It bears no relation to any of the phenomena previously encountered in electrical science. When an anode was first placed in a vacuum with a hot cathode, it was the first time in history that current flowed across an open space without the usual accompanying arc or ionization of gas. As such, it was found that the currents and voltages in a thermionic circuit do not act in accordance with the laws heretofore formulated. For example, the currents and voltages (a-c components) in the plate circuit of a vacuum tube do not vary in precise accordance with Ohm's law at all times. Consequently, new methods of computation were devised utilizing graphs by means of which a vacuum tube's performance can be predicted for given circuit conditions. Such graphs are necessary adjuncts when any calculations are to be made concerning a vacuum tube circuit. They are empirical in origin, that is, they are obtained by actual measurement of circuit constants under varying conditions and are called **characteristic curves.** Characteristic curves for specific tubes are customarily provided by the tube manufacturers.

There are three variables that affect the plate current in a vacuum tube: the plate voltage, the grid voltage, and the heater voltage. Under ordinary conditions, the cathode is operated at or near saturation point. The heater voltage is maintained at a specific fixed point at which maximum operating efficiency is obtained, so that this voltage may, therefore,

be considered a fixed value in actual practice. With the heater voltage a constant, the plate current is dependent upon two remaining variables—the grid voltage and the plate voltage. The manner in which these variables affect the plate current is a function of the tube characteristics. When the variations in plate current resulting from variations of either

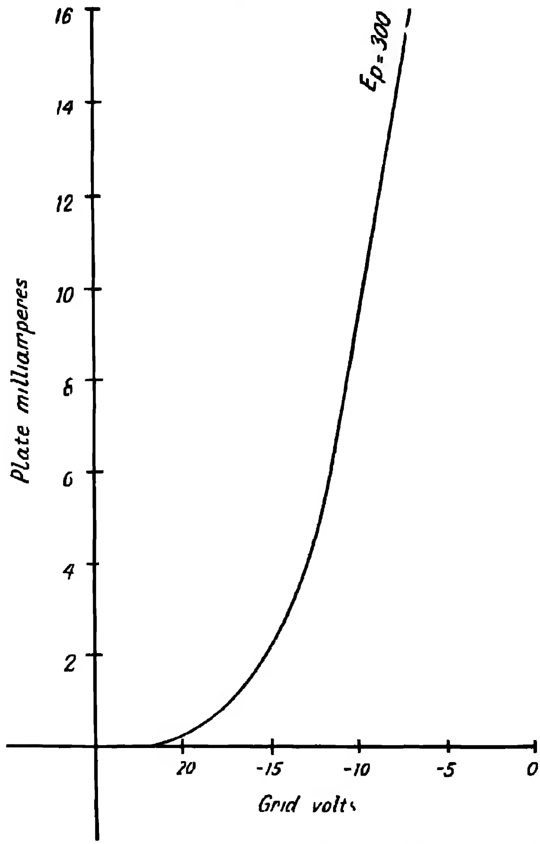



FIG. 110 An $E_c$, $I_p$ curve for a type 6C5 tube when $E_p$ 300 v

the grid voltage or the plate voltage are plotted in graphs, the resultant curves obtained are characteristic curves.

Grid curves are the characteristic curves obtained when plate current is plotted against grid voltage with the plate voltage constant Figure 110 illustrates a single grid curve for a triode tube, in which the grid voltage is plotted on the $X$ axis and plate current on the $Y$ axis. Such a curve is not complete unless the value of plate voltage at which it was computed is known. The curve of Fig 110 gives the value of plate current for various values of grid voltage with the plate voltage fixed at 300 v

for a type 6C5 tube. Since the amount of information obtainable from a single curve of this type is limited by the condition of fixed plate voltage, standard practice is to provide a number of such curves for a given tube. Such a group of curves gives the relation of plate current to grid voltage for a number of different values of plate voltage and is called a **family**

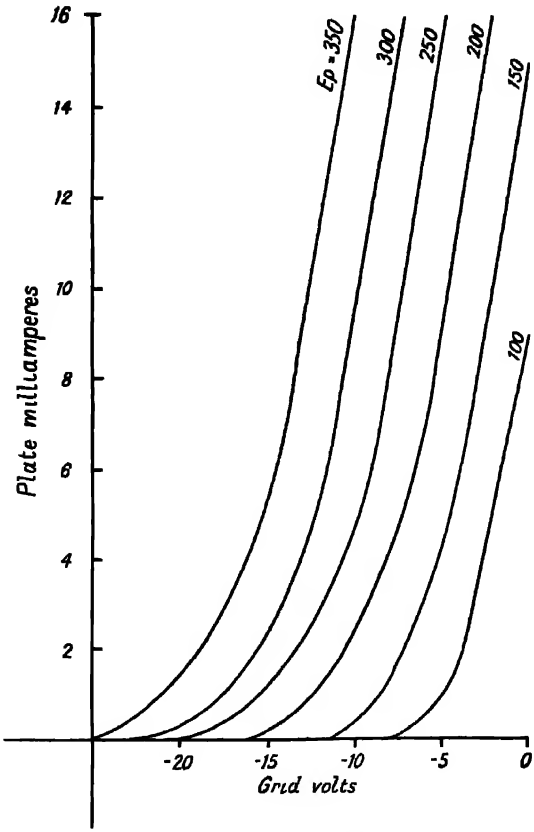
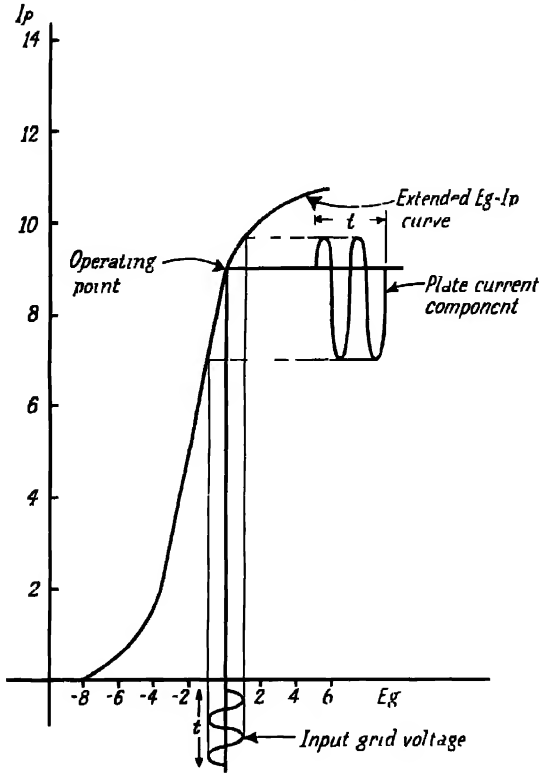



FIG. 111 A family of $E_g I_p$ curves for a 6C5 tube

**of grid curves.** Figure 111 illustrates a family of grid curves for the type 6C5 tube. Grid curves are also commonly called "$E_g$ $I_p$ curves."

Some important and interesting facts may be ascertained from grid curves: From the steepness of the curve the amplification possible with the tube may be obtained, from the degree of linearity of the curve, the amount of distortion is learned, and the proper portion of the curve over which to operate the tube is thus deduced. In short, the proper values of grid and plate voltage necessary for maximum output and minimum distortion may be ascertained from inspection of a grid family of curves.

Figure 112(a) illustrates a single extended grid curve for a 6C5 tube at a plate voltage of 100 v. Suppose that the tube is to be used as an amplifier of pure sine-wave alternating current with a peak value of 1 v. This alternating voltage can be projected on the graph, as shown in Fig. 112(a). It is at once apparent that considerable mutilation of the plate-

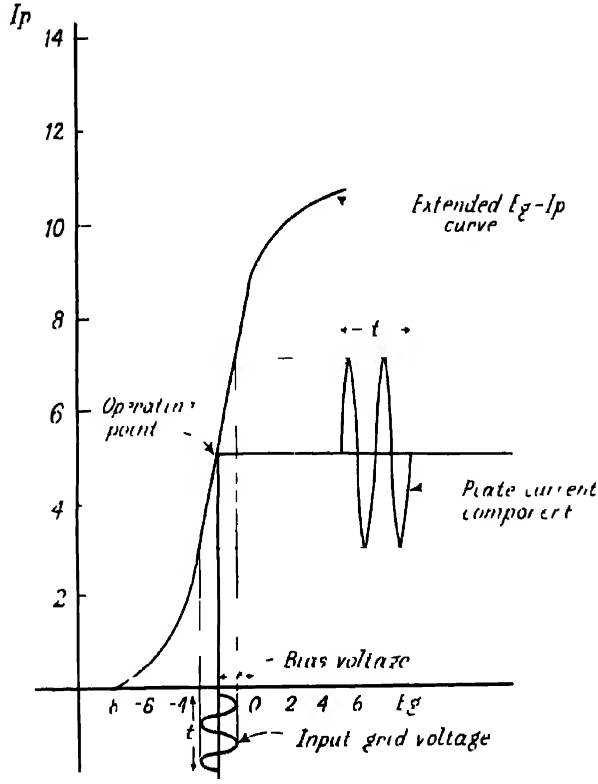



(a)

FIG. 112(a). Input and output wave forms for a 6C5 tube operating as an amplifier without bias voltage. Notice the output distortion due to the curvature of the characteristic.

current wave form occurs The plate-current component is not identical with the input signal voltage, that is, distortion has occurred because the tube was operated at the upper bend of the grid curve. There is a lack of linearity because the grid is positive on every other alternation of the signal voltage. The plate current, therefore, does not increase as it should, because part of the electrons, instead of going to the plate of the tube, are attracted to the positive grid, causing a flow of current in the grid circuit. Obviously, in order to minimize distortion of this nature, the

tube should be operated over a linear portion of the grid curve. In order to accomplish this, a fixed value of d c voltage is impressed across the grid-cathode circuit of the tube, which can be done by inserting a battery in this circuit. (Several methods of obtaining grid bias are discussed later in the chapter.) For the 6C5 tube under discussion, a 2-v battery




(b)

Fig. 112(b) Input and output wave forms for 6C5 tube operating with − 2 v. bias. Notice the symmetry of the output wave form and the over all amplification

would serve adequately. This brings the operating point of the tube into the approximate center of the linear portion of the curve. The impressed 1-v alternating emf would subsequently cause the grid voltage to vary from − 1 to − 3 v. The operating point would never be off the linear portion of the curve. The projected voltage and resultant plate current for this condition are illustrated in Fig. 112(b). The d-c voltage used for this purpose is called the **bias voltage**; and a tube operating under these conditions is said to be **biased**. If a battery is used to provide

the bias voltage, it is called a C battery to distinguish it from the A and B batteries used in other parts of the circuit.

**Plate curves** are the characteristic curves obtained when the plate current is plotted against plate voltage with the grid voltage constant. In common with the grid curves, plate curves are also customarily plotted in families, each separate curve expresses the relation of plate current

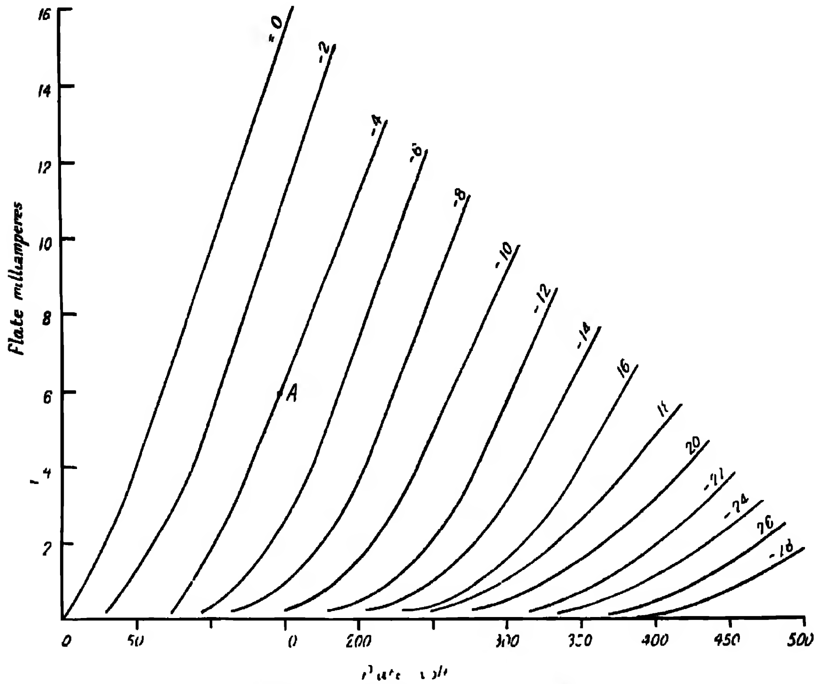



FIG. 113 A family of $E_p$, $I_p$ curves for a type 6C5 tube

to plate voltage for a different value of grid voltage. Thus a maximum amount of information is derived from a single graph.

From a **family of plate curves,** all the constants of the tube may be calculated, and the performance of the tube when connected into a circuit of known electrical constants can be foretold from these curves. A family of plate curves for the 6C5 tube is shown in Fig. 113. The application of plate curves is discussed in a later part of this chapter.

## *AMPLIFIER CLASSIFICATIONS*

Although it would appear that every amplifier tube should be operated at the center of the linear portion of its grid curve, it is often desirable to operate a tube at other points of the curve. The advantages of such

operation will become evident as this discussion progresses. In order to distinguish between amplifiers operated at different points of the grid curves. they have been divided into three broad classifications: class A amplifiers, class B amplifiers, and class C amplifiers.

**Class A Amplification.** A class A amplifier is one in which the grid bias and the exciting grid voltage (signal voltage) are such that the plate current through the tube flows at all times. Such an amplifier operates at the center of the linear portion of its grid curve, and the plate output wave form is essentially the same as that of the exciting grid voltage. For normal class A operation, the grid must not be allowed to go positive on excitation peaks, and the plate current must not fall low enough at its minimum to cause distortion resulting from operating on the lower curved portion of the characteristic. Class A amplification is accomplished by negatively biasing the tube by an amount that will keep the operating point at the center of the linear portion of the $E_g$-$I_p$ curve. For a specific class A application, a tube is chosen whose grid characteristic curve is long enough to prevent the signal voltage from traversing either the upper or the lower curved portion of the characteristic. The grid curve of a tube operating class A is illustrated in Fig. 114 with the input and output wave forms.

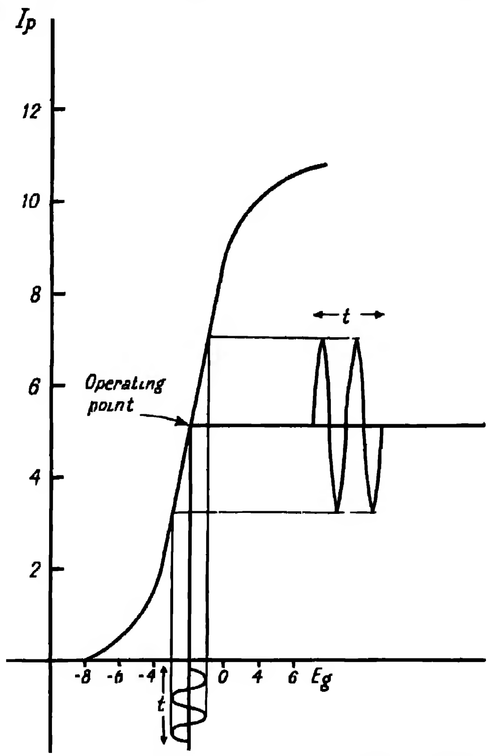



FIG. 114 Input and output wave forms for class A operation

Class A amplification is used wherever it is desired to obtain good fidelity and where efficiency and gain per stage are not of paramount importance.

**Class B Amplification.** A class B amplifier is one in which the negative grid bias is kept sufficiently high so that very little or no plate current flows in the absence of exciting signal voltage. The value of bias is such that any positive signal making the grid less negative will allow plate

current to flow. The point on the grid curve at which such operation occurs is called the **cutoff point,** since plate current ceases to flow, or is cut off, when this value of bias is reached. In such an amplifier, plate current flows during one half of the cycle when an exciting grid voltage is present. The ideal class B amplifier is one in which the alternating

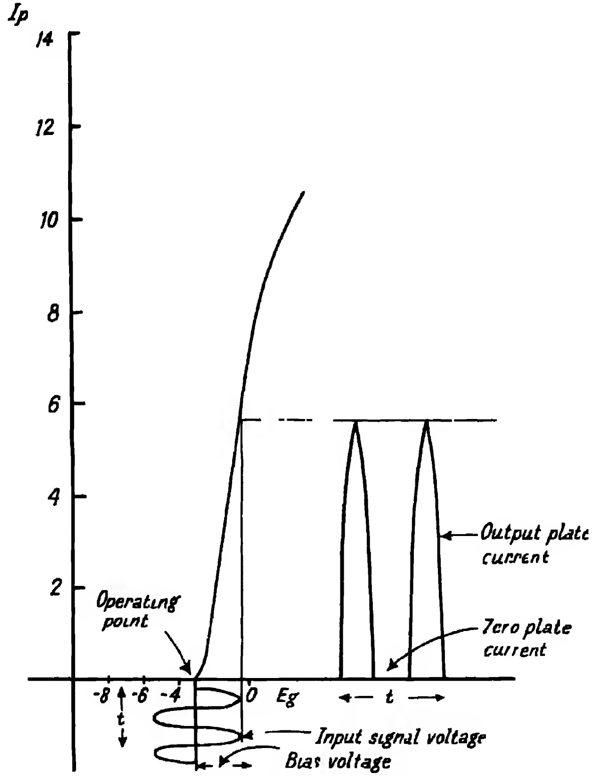



FIG. 115 Input and output wave forms for class B operation (single tube).

component of plate current is an exact replica of the alternating grid voltage for the half cycle during which plate current flows. The characteristics of class B amplifiers are medium efficiency and output. Figure 115 illustrates the grid curve of a class B amplifier with the input and output wave forms.

It is at once apparent that the output of a single tube operating class B is by no means identical with the input wave form. The alternating component of the plate current consists of a series of pulses corresponding to every other half cycle of input voltage. Class B amplifiers, therefore, utilize two tubes per stage operating push pull. A push-pull circuit (described in detail in Chap. XII) is an arrangement wherein the same

signal voltage is applied to the input circuits of two class B tubes in such a manner that each tube functions during the half cycle in which the other tube is idle. The outputs of the two tubes are coupled 180° out of phase to a common load, thus completing the original wave form. A schematic diagram of a typical push-pull amplifier is shown in Fig. 116. The input and output wave forms for such an amplifier operating as class B are shown in Fig. 117.

Class B amplifiers are used where very high output of good quality utilizing fairly small tubes operating at relatively low plate voltage is desired. Unusual over-all economy of power consumption is possible because the plate current is low (zero under ideal conditions) when no signal is applied to the grid. By designing class B amplifier tubes with a

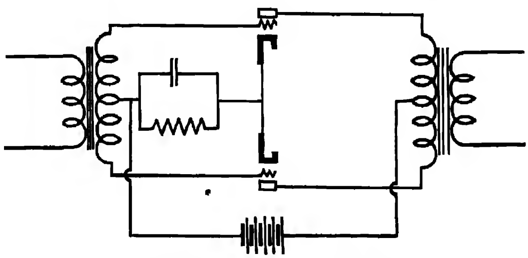


FIG. 116. A typical class B push pull amplifier circuit

sufficiently high amplification factor, it is possible to operate them with plate-current cutoff occurring at or near zero grid voltage on the grid curve. It is therefore possible to dispense with C batteries and bias resistors.

**Class C Amplification.** A class C amplifier is one in which the grid bias is kept at a value appreciably beyond the cutoff value. The plate current in each tube is zero when no exciting grid voltage is present. Plate current flows in each tube for appreciably *less* than one-half of each cycle when an exciting grid voltage is present. Figure 118 illustrates the input and output wave form for a single tube of a class C amplifier. Class C amplifiers are characterized by high plate-circuit efficiency and comparatively high distortion.

Class C amplifiers find application where high plate-circuit efficiency is a paramount requirement and where departures from linearity between input and output are permissible. The inherent distortion in this type of amplification precludes its use for any kind of voice-modulated signal or for audio frequency. Class C amplifiers are not used in receivers but find their greatest application in transmitters as r-f amplifiers. They are used in telegraph transmitters and in telephone (broadcast) transmitters in stages preceding the modulated stage.

**Intermediate Classifications.** It is often convenient to have terms to

identify amplifier services when the operating conditions are intermediate to those of classes A and B or to those of classes B and C. Such operating conditions have been classified as class AB and class BC, respectively.

A class AB amplifier partakes of the characteristics of both the class A and the class B amplifiers, although not adhering strictly to either.

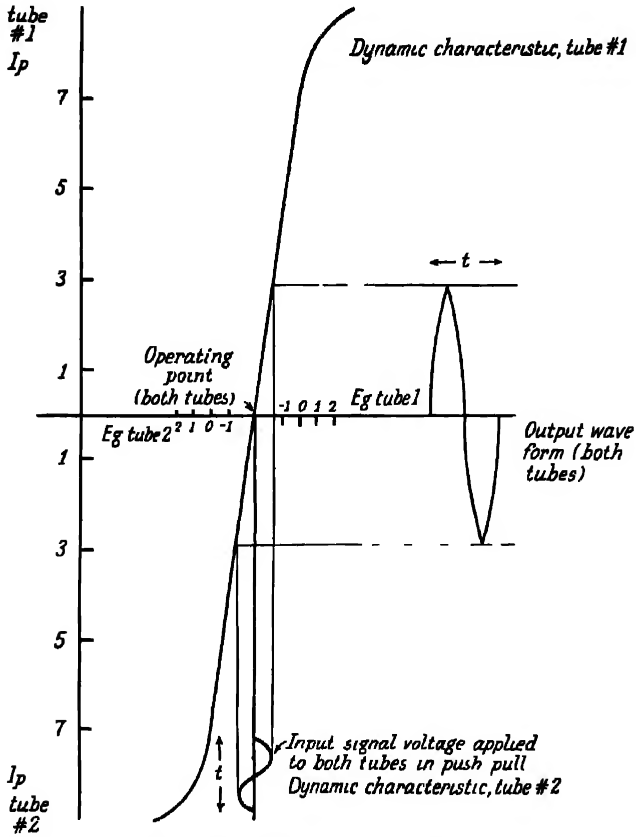



FIG. 117. Input and output wave forms for two tubes operating class B in a push-pull amplifier.

The grid bias and exciting grid voltage in a class AB amplifier are kept at such a value that plate current flows during appreciably more than 180° but less than 360° of the cycle. The no-signal plate current and resultant plate dissipation can be made substantially less with this class of amplifier than is possible with the class A. The class AB amplifier retains the advantages of class B operation and at the same time overcomes the chief objection to class B amplifiers—the distortion attendant with low values of signal voltage. A class AB amplifier tends to operate

class A with low values of signal voltage and class B with high signal voltages. This classification is widely used wherever it is desired to obtain high power output with a minimum of distortion.

A class BC amplifier is one in which the grid bias and exciting grid voltage are kept at such a value that plate current flows during a considerable portion of the cycle but less than 180°. The efficiency and output of a class BC amplifier are intermediate to those of the class B and C. Class BC amplifiers are not in general use.

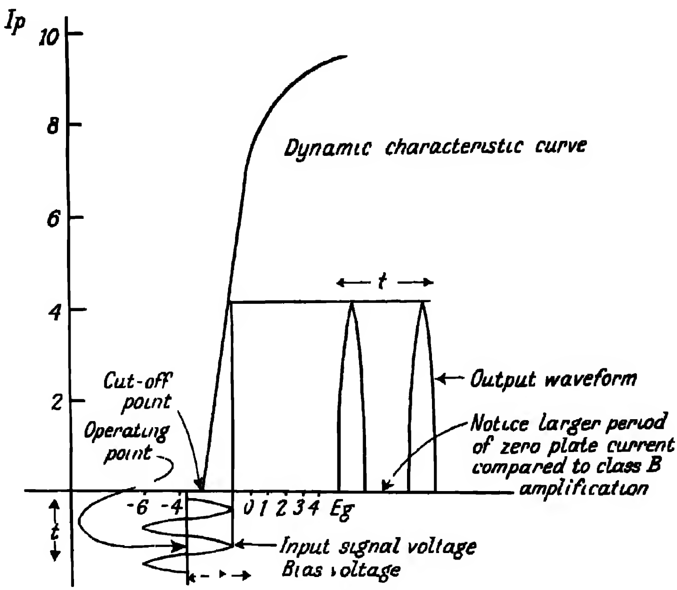



Fig. 118. Input and output wave forms for class C operation (single tube)

The exciting grid voltage applied to *any* of the various classes of amplifiers can be increased to the point where the grid becomes positive for a portion of the cycle. Under this condition, current flows in the grid circuit and power is, therefore, absorbed by this circuit. In order to facilitate the handling of such circuits, the condition of amplifier operation may be further identified by the addition of a suffix to the classification. The subscript suffix 1 added to the classification of an amplifier indicates that grid current does *not* flow during any part of the input cycle. The subscript suffix 2 indicates that grid current flows during some portion of the input cycle. Thus, if a class AB amplifier is operated under conditions that permit the grid to become positive during a portion of the cycle, it should be classified as a class $AB_2$ amplifier. If the grid remains negative throughout the entire cycle, it should be classified as a class $AB_1$ amplifier. The latter classification is often referred to as "class AB prime."

## *GENERAL AMPLIFIER CONSIDERATIONS*

Vacuum-tube amplifiers can be grouped into two divisions so far as their application in practical circuits is concerned. An amplifier can be used to amplify *power* or simply to amplify *voltage*. A tube whose grid is negative throughout the entire cycle of exciting grid voltage has no current flowing in the grid circuit at any time. No power is required by the input circuit of such a tube; all that is necessary to operate the tube as an amplifier is to apply a *voltage* across its grid-cathode (input) circuit. Since the grid is always negative, no electrons from the cathode are attracted to it. The grid-cathode circuit is essentially open, therefore, so far as the input signal is concerned. The amplifier *preceding* such a circuit need only supply *voltage* and is consequently called a **voltage amplifier.**

If a tube is so operated that its grid is positive during some portion of the input cycle, current flows in the grid circuit during this period. Part of the electrons emitted from the cathode are attracted to the positive grid. These electrons return to the cathode through the grid circuit and hence constitute a flow of current in this circuit. The d c bias voltage has no part in causing this current flow, since the bias voltage on the grid is negative. The current flow is caused by that portion of the exciting grid voltage that causes the grid to go positive. The power expended in the grid circuit, therefore, is the product of this portion of the excitation voltage and the current that it causes. Since the power expended in this manner is a function of the excitation voltage, it must be supplied from the excitation source. Consequently, the amplifier preceding such a circuit must deliver sufficient power to drive the circuit properly. Such an amplifier is called a **power amplifier.**

Either a voltage amplifier or a power amplifier may be operated under any of the classifications previously discussed The restrictions limiting the choice of the proper class of amplification are the permissible amount of distortion, the available excitation, the magnitude of the output required, the plate dissipation, and the plate circuit efficiency.

The power delivered from the output circuit of a vacuum tube originates in the d-c plate supply. The ratio of the power usefully delivered to the output circuit and the power supplied by the d-c source is an indication of the plate-circuit efficiency of the tube. The power not usefully expended in the output is expended in heating the plate of the tube and is called the **plate dissipation.** The amount of power that can be transferred from a tube circuit to a load circuit is dependent upon two factors. The first factor is the amount of power available in the plate circuit, and the second factor is the ratio of the load resistance to the internal (plate) resistance of the tube.

The amount of power available in the plate circuit of any tube is a

function of the voltage applied to the plate and the current flowing in the plate circuit. Hence, it is apparent that in order to obtain high power output from a tube, the tube must be operated at a high plate voltage, or at a high plate current, or both. In the smaller power tubes, such as are used in radio receivers, the plate voltage is limited to 250 or 300 v by practical requirements. Large power output is therefore obtained by increasing the plate current, which necessitates the use of a high-emission cathode. The plate in such tubes must be placed in closer proximity to the cathode in order that the comparatively small plate charge may neutralize properly the large space charge of such an emitter. Consequently, tubes of this type are characterized by low amplification factors.

In large power tubes, such as those utilized for transmission, the plates are designed to operate at very high plate voltages and relatively small plate currents. Higher plate-circuit efficiency is possible with such tubes. Less excitation is also required because the low plate current (low cathode emission) permits higher amplification factors. It is evident, therefore, that the type of tube best suited for power amplification differs from the type of tube best adapted for voltage amplification. For the latter application, the power output need not be considered at all.

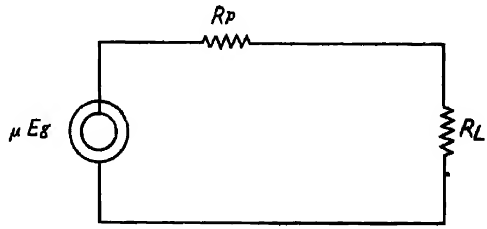


FIG. 119 Equivalent circuit for a vacuum tube working into a load.

The ratio of load resistance to internal resistance is a factor that affects the transfer of energy (either voltage or power) in any kind of circuit. Whenever a source of energy is connected to a load circuit, maximum power is transferred to the load only under certain optimum conditions, and maximum voltage is impressed across the load only under certain other optimum conditions. A vacuum-tube amplifier may be considered as a source of energy having an internal resistance $R_p$ and generating a voltage $\mu E_g$. This is shown in the equivalent circuit in Fig. 119, where $R_L$ is the resistance of the load circuit.

It can be shown that in such a network maximum *voltage* is developed across the load resistance when this resistance is made infinitely high. Since this is, of course, an impossibility, maximum output voltage is attained by making the load resistance as high as possible compared with the plate resistance of the tube. This may best be appreciated by assuming a theoretical voltage $\mu E_g$ developed by a tube of given plate resistance $R_p$ and calculating the voltage output for various values of load resistance. Thus, if $\mu E_g$ 100 v for a tube having a plate resistance

of 1,000 ohms and a 500-ohm load is coupled to the tube, the current in Fig. 119 will be

$$I = \frac{E}{R_p + R_L} = \frac{100}{1{,}500} = 0.066 \text{ amp.} \tag{9}$$

The voltage developed across $R_L$ will be

$$E = I \cdot R = 0.066 \cdot 500 = 33 \text{ v.} \tag{10}$$

If $R_L$ is made 1,000 ohms, the total circuit current will be

$$I = \frac{E}{R_p + R_L} = \frac{100}{2{,}000} = 0.05 \text{ amp.} \tag{11}$$

The voltage developed across $R_L$ will be

$$E = I \cdot R = 0.05 \cdot 1{,}000 = 50 \text{ v.} \tag{12}$$

If $R_L$ is made 2,000 ohms, the total circuit current becomes

$$I = \frac{E}{R_p + R_L} = \frac{100}{3{,}000} = 0.033 \text{ amp.} \tag{13}$$

The voltage developed across $R_L$ becomes

$$E = I \cdot R = 0.033 \cdot 2{,}000 = 66 \text{ v.} \tag{14}$$

Obviously, as $R_L$ is increased, the voltage developed across it is increased. Therefore, the ratio of load resistance to plate resistance in a voltage amplifier should be made as high as possible in order to attain maximum voltage output. In actual practice a load resistance three or four times the plate resistance will be found to give satisfactory results. In r-f amplifiers, the phenomenon of parallel resonance is utilized to good advantage in this respect. The load is taken from a parallel $L$ and $C$ circuit, which is operated at resonance and thus offers a very large impedance to the plate current: for that reason such circuits function very efficiently as voltage amplifiers.

Maximum *power* is developed in the load resistance when this resistance is exactly *equal* to the plate resistance. Thus, if we utilize the figures of the above example, the power developed in the load resistance of Fig. 119 when $R_L$ is 500 ohms is

$$P = I^2R_L = 0.066^2 \cdot 500 = 2.178 \text{ w.} \tag{15}$$

When $R_L$ is increased to 1,000 ohms, the circuit current [taken from Eq. (11)] is 0.05 amp and the power developed in the load resistance becomes

$$P = I^2R = 0.05^2 \cdot 1{,}000 = 2.5 \text{ w.} \tag{16}$$

When $R_L$ is made 1,500 ohms, the circuit current is 0.04 amp and the power in $R_L$ becomes

$$P = I^2R = 0.04^2 \cdot 1{,}500 = 2.4 \text{ w.} \tag{17}$$

When $R_L$ is made 2,000 ohms, the circuit current is 0.033 amp and the power in the load becomes:

$$P = I^2R = 0.033^2 \cdot 2{,}000 = 2.178 \text{ w.} \tag{18}$$

Thus, for values of load resistance higher or lower than the plate resistance, the power output falls off. Maximum power output is obtained, therefore, when the load resistance equals the plate resistance. It will be shown subsequently that although maximum power output is obtained under these conditions, maximum *undistorted* power is obtained when the load resistance is made approximately twice the plate resistance of the tube.

For a tube of known constants, the power output for varying conditions of load resistance can be computed if the excitation voltage available is known. Thus,

$$P = I^2R_L, \tag{19}$$

where $P$ = power developed in the load resistance,
$I$ = plate current in the tube:
$R_L$ = resistance of load.

$$I = \frac{\mu E_g}{R_p + R_L}. \tag{20}$$

Substituting the equivalent value of $I$ from Eq. (20) in Eq. (19) the latter expression becomes

$$P = \frac{\mu^2 E_g{}^2 R_L}{(R_p + R_L)^2}. \tag{21}$$

Substitutions may be made directly in Eq. (21) to determine the power output for any value of load resistance when the excitation voltage $E_g$ is known.

Equation (21) can be further simplified for the condition of maximum power output, that is, when $R_p$ and $R_L$ are equal. Substituting for $R_L$ its equal value $R_p$, Eq. (21) becomes

$$P = \frac{\mu^2 E_g{}^2 R_p}{(2R_p)^2} = \frac{\mu^2 E_g{}^2 R_p}{4R_p{}^2}, \tag{22}$$

and

$$P = \frac{\mu^2 E_g{}^2}{4R_p}. \tag{23}$$

It should be remembered that the values of excitation voltage $E_g$ substituted in Eqs. (21) and (23) are rms, or effective, values.

**Load-Line Curves.** Figure 113 illustrates a plate family of curves for the type 6C5 tube. Such curves, in common with all other curves heretofore discussed, are called **static curves,** since they give the tube's performance for a number of *fixed*, or static, conditions. In order to predict

accurately the performance of a tube under actual operating conditions, it is necessary to utilize curves that represent this performance under actual circuit conditions. Such curves are called **dynamic curves** because they give the performance of the tube in the presence of a continually varying excitation voltage.

Thus, suppose it is desired to operate the 6C5 tube mentioned above as a power amplifier with a plate voltage of 150 v and a negative bias of $-4$ v. The operating point on the plate family of Fig. 113 will be fixed at point $A$ on the $E_g = -4$ curve. Under actual circuit conditions, however, there will be a load resistance in the plate circuit of the tube. The plate current will not be 6 ma, as obtained from Fig. 111, but will take a value determined by the limiting resistance of $R_p + R_L$. The operating point will therefore be shifted from point $A$ to a point lower down on the curve. If an alternating signal voltage is now impressed across the input circuit of the tube, the negative grid voltage will be either added to or subtracted from by an amount equal to the peak value of signal voltage, and dependent upon whether the signal alternation is positive or negative at the instant. At the moments when the grid is less negative, the plate current will *increase*. Hence, the voltage drop across the load resistance $R_L$ will be greater, with the result that the actual voltage impressed on the plate of the tube will be less. Conversely, when the grid is *more* negative on the other half of the input cycle, plate current will *decrease*. The voltage drop across the load resistance will therefore be less and the voltage on the plate of the tube greater. It is apparent that the presence of a load resistance in the plate circuit materially alters the static characteristic curves of a tube. Dynamic plate-characteristic curves are $E_p$-$I_p$ curves that have been plotted for a number of different values of load resistance for a fixed value of grid voltage. It has been determined that inserting a load in the plate circuit of a tube causes its characteristic to become flatter with a longer linear portion. In order to predict the performance of a tube completely, it would be desirable to have available a number of $E_p$-$I_p$ dynamic families for a number of different values of grid voltage. The necessity for handling so cumbersome an amount of graphs is obviated by a simple method of computing the performance of a tube with a given load resistance, which is done graphically by calculating the *load line* of a tube.

Assume that it is desired to calculate the performance of a type 45 tube with a load resistance of 4,000 ohms operating at a plate voltage of 250 v with a grid bias of $-50$ v. Figure 120 illustrates a static plate family for this tube. Point $P$ is the proposed operating point. The procedure is as follows. A straight line is drawn intersecting the $E_p$ axis and the $I_p$ axis. The slope of this line is the reciprocal of the load resistance and is therefore easily determined by taking any voltage,

such as 100 v, and ascertaining by Ohm's law the current in a resistance equal to the load resistance as 4,000 ohms. Thus,

$$I = \frac{E}{R} = \frac{100}{4{,}000} = 0.025 \text{ amp} = 25 \text{ ma.} \tag{24}$$

The line in question must therefore intersect the $E_p$ axis at the 100-v

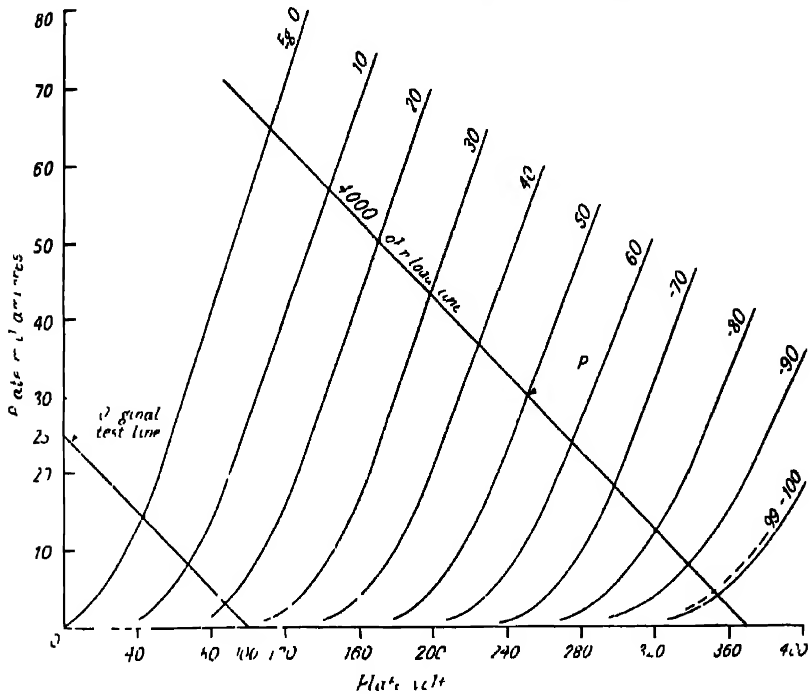



FIG. 120. Plate family for a type 45 tube illustrating the method of obtaining a load line

point and must intersect the $I_p$ axis at the 25 ma point, as shown in Fig. 120.

Through point $P$ (the operating point) draw a line parallel to the line just constructed above. The line through $P$ is called the **load line,** since it has the proper slope for the load in question (4,000 ohms) through the desired operating point.

All the information essential to the proper operating of this tube with this load resistance may now be obtained from the load line. The load line is actually the locus of the operating point for a given load resistance under fixed conditions of plate voltage and grid bias. Thus, since it is known that distortion results from operating the tube on the curved portion of its static characteristic, the lowest plate current that can be

tolerated for the 45 tube in question will be approximately 4.5 ma, or the value when $E_g = -99$ v. The peak a-c voltage that should be applied to the grid of this tube, therefore, is 49 v. Such an input signal voltage will cause the grid voltage to vary from $-1$ to $-99$ v with a resultant plate-current variation from 64 to 4.5 ma. The resultant variation of the voltage at the plate is found from the load line to be from 112 to 345 v. The necessary power supply, or B battery voltage, is determined by the intersection of the load line with the coordinate axis, which in this case yields 370 v.

The power output for a given tube and load resistance may be calculated from data obtained from the load line by means of the following formula:

$$\text{power output} = \frac{(E_{\max} - E_{\min}) \times (I_{p,\max} - I_{p,\min})}{8}, \tag{25}$$

where $E_{\max}$ = maximum value of plate voltage,
$E_{\min}$ = minimum value of plate voltage;
$I_{p,\max}$ = maximum value of plate current;
$I_{p,\min}$ = minimum value of plate current.

Thus, for the type 45 tube under the conditions discussed above,

$$\text{power output} = \frac{(345 - 112) \times (0.064 - 0.0045)}{8}, \tag{26}$$

$$\text{power output} = \frac{(233)(0.0595)}{8} = \frac{13.85}{8}, \tag{27}$$

$$\text{power output} = 1.73 \text{ w} = 1{,}730 \text{ mw}. \tag{28}$$

Distortion in vacuum tube circuits is evidenced by unsymmetrical wave forms in the output. The energy represented by such distortion is present in the plate circuit in the form of harmonics of the operating frequency. Harmonics are multiples of the fundamental frequency and are discussed in detail in a later chapter. By far the largest percentage of distortion occurring in amplifiers with one tube per stage is in the form of second harmonic distortion. The higher orders of harmonics generated (third, fourth, fifth, and so on) contain progressively less energy and can usually be neglected. Under ordinary conditions, third-harmonic energy is always less than second-harmonic energy, fourth-harmonic energy less than third harmonic, and so on. Thus, if the second-harmonic content is kept down to an unobjectionable value, the higher-order harmonics may be neglected. The percentage of second-harmonic distortion present in an amplifier can be computed from the load-line curve by the formula

$$\left(\begin{array}{c}\text{percentage second} \\ \text{harmonic distortion}\end{array}\right) = \frac{\frac{1}{2}(I_{p,\max} + I_{p,\min}) - I_b}{(I_{p,\max} - I_{p,\min})} \cdot 100, \tag{29}$$

where $I_{p,\,max}$ = maximum value of plate current;
$I_{p,\,min}$ = minimum value of plate current;
$I_n$ = normal plate current (no signal voltage).

Thus, for the type of 45 tube under the conditions discussed above, the percentage of second-harmonic distortion becomes

$$\%\ \text{second-harmonic distortion} = \frac{\frac{1}{2}(0.064 + 0.0045) - 0.030}{(0.064 - 0.0045)} \cdot 100, \quad (30)$$

$$\%\ \text{second-harmonic distortion} = \frac{0.0342 - 0.030}{0.0595} \cdot 100, \quad (31)$$

$$\%\ \text{second-harmonic distortion} = \frac{0.0042}{0.0595} \cdot 100 = 0.07(100) = 7. \quad (32)$$

Thus, a type 45 tube working into a 4,000-ohm load with 250 v on the plate, −50 v on the grid, and a peak signal-voltage input of 49 v will produce a power output of 1,730 mw. The percentage of distortion with this operation will be 7 per cent.

The percentage of distortion in such an amplifier is a function of load resistance. If a number of load lines are plotted for a given tube, a desirable load resistance may be obtained which will reduce the harmonic distortion to an unobjectionable value without too greatly decreasing the power output. Standard practice is to utilize a load resistance such that the distortion does not exceed 5 per cent, a value that experience has shown to be permissible. Several approximations of load resistance may be necessary to obtain the optimum value for the operating conditions chosen. It has been determined that maximum *undistorted* power output is obtained when the load resistance is approximately twice the plate resistance of the tube.

## *MULTIELEMENT VACUUM TUBES*

**The Tetrode.** The electrodes that comprise the elements of a vacuum tube, like any other electrodes, possess electrostatic capacitance. When any two conductors are placed in proximity to each other, they can be said to form a capacitor. The electrodes of a tube plate, grid, cathode, and so on, possess, owing to their relative proximity, a small amount of capacitance in conjunction with each other. The capacitive reactance is high because of the small physical size of tube electrodes and at low frequencies can often be neglected. At high frequencies, however, this reactance often becomes fairly low, since capacitive reactance varies inversely with the frequency. This is especially true of tubes with high amplification factors, since such tubes have their electrodes spaced more closely and have smaller mesh openings on the grid, thus increasing the area and, hence, the capacitance of the latter electrode. A low value of

interelectrode capacitive reactance in a tube, especially between control grid and plate, is undesirable, since it provides a path for the a-c component of plate current to feed back into the grid circuit. This component is then reamplified by the tube with resulting distortion, often causing the tube to oscillate (become a self-sustaining generator of alternating emf). Under such conditions, steps must be taken to neutralize this feedback in order to prevent these undesirable effects. The procedure is discussed in a later chapter devoted to oscillators and neutralization.

The above condition was prevented by the insertion of an electrostatic shield between the control grid and plate of the vacuum tube. In order to avert undue interference with the plate electron flow, this shield is constructed of wire mesh and is often called the **screen grid.** The screen grid is kept at ground potential with respect to a-c voltages by means of an external capacitor of low reactance connecting it to the cathode. Since the plate is effectively shielded from the other electrodes by the screen grid, its positive charge is much less effective in attracting electrons from the grid. A positive d-c potential is applied to the screen and thus partially takes over the job of attracting electrons from the cathode. When the electrons reach the screen grid most of them have sufficient velocity to continue on through the mesh openings to the plate whose greater positive charge is more effective at this point. A small portion of them, however, are attracted to the screen grid and constitute a flow of current in the screen-grid circuit.

The space charge about the cathode of a tube is one of the factors that limit the flow of current to the plate. The screen grid, owing to its positive potential and closer proximity to the cathode, greatly reduces the space charge. The control grid, therefore, has a much greater controlling effect on the plate current, with the result that the tetrode, or screen-grid tube, has a much greater amplification factor than the triode. In addition, the screen grid exercises most of the attraction for the electrons, so that a given change in plate voltage causes a much smaller change in plate current than with a triode tube. Consequently, the plate resistance of a screen-grid tube, according to Eq. (3), is much higher than that of a triode. This latter characteristic makes the tetrode particularly adaptable for use in tuned amplifier circuits where it is necessary to match the plate resistance to the high impedance of a parallel resonant circuit.

The electrons that arrive at the plate of a tube often have sufficient velocity to knock off, or dislodge, other electrons from the plate. These electrons are called **secondary electrons** and this phenomenon is known as **secondary emission.** The screen grid, because of its positive potential, attracts the electrons freed by secondary emission. If the plate voltage becomes lower than the screen voltage, secondary emission is particularly pronounced, and the plate current is seriously decreased. The plate swing is therefore limited by the effects of secondary emission.

**The Pentode.** In order to remove the limitation of plate current imposed by secondary emission a fifth electrode, called the **suppressor grid,** is inserted in the vacuum tube The suppressor is placed between screen grid and plate and is connected to the cathode The suppressor is therefore negative with respect to the plate and repels the secondary

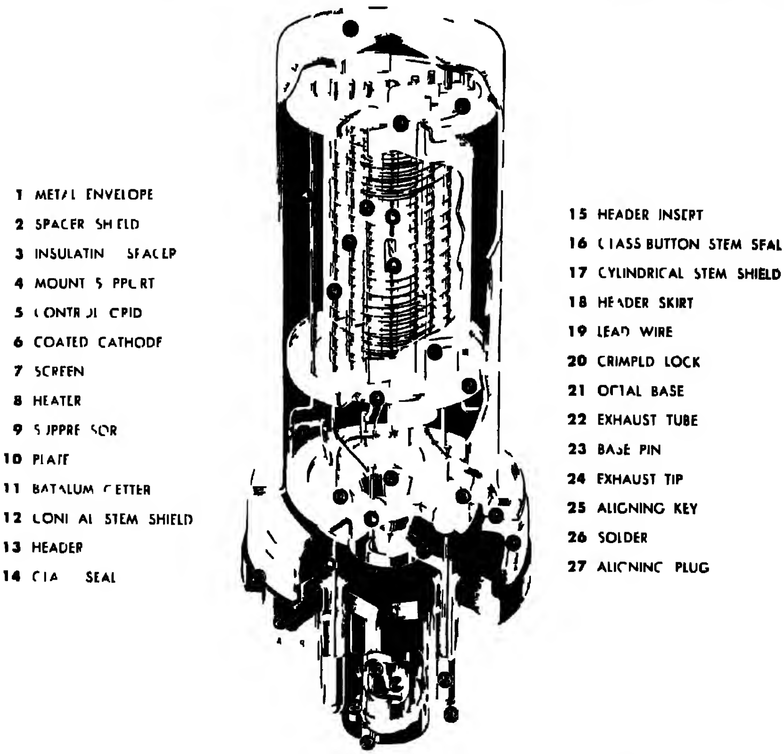



Fig. 121 Structure of single-ended metal tube (*Courtesy of RCA Inc.*)

electrons back to the plate This electrode receives its name from its action in suppressing the flow of secondary electrons In power pentodes, the suppressor permits a large power output with high gain because the plate voltage swing can be made very large In voltage amplifiers the pentode is capable of high voltage amplification at moderate values of plate voltage

## *QUESTIONS AND PROBLEMS**

**1.** Describe the electrical characteristics of the pentode tetrode and triode on a comparative basis

* These questions and problems are taken from the F C C Study Guide for Commercial Radio Operator Examinations

**2.** What is mutual conductance? Transconductance?

**3.** What is secondary emission?

**4.** Describe the characteristics of a vacuum tube operating as a class C amplifier.

**5.** During what portion of the excitation-voltage cycle does plate current flow when a tube is used as a class B amplifier?

**6.** Does a properly operated class A audio amplifier produce serious modification of the input wave form?

**7.** A triode transmitting tube, operating with plate voltage of 1,250 v, has filament voltage of 10, filament current of 3.25, and plate current of 150 ma. The amplification factor is 25. What value of control grid bias must be used for operation as a class C stage?

**8.** Discuss the advantages and disadvantages of operating an amplifier as a class C stage

**9.** What circuit and vacuum-tube factors influence the voltage gain of a triode a-f amplifier stage?

**10.** What is electron emission?

# Chapter XI

# THE VACUUM-TUBE OSCILLATOR

One of the most important applications of the vacuum tube is as a generator of alternating current. A vacuum tube can be made to generate alternating currents of frequencies ranging from the very lowest usable commercial frequencies to frequencies on the order of *several hundred million cycles per second.* It is possible to generate higher-frequency alternating currents by means of vacuum tubes than by any other means.

Radio is primarily the application of h-f alternating currents. Because of its superiority as a generator of such currents, the vacuum tube has

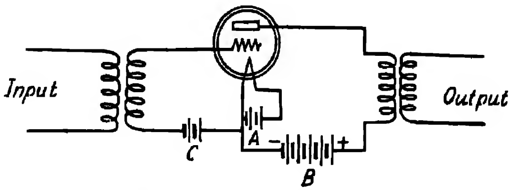



FIG. 122. An elementary vacuum tube amplifier.

superseded practically every other means as an h-f generator in radio work.

Because alternating currents are basically oscillating currents, a vacuum tube is called an **oscillator** when used to generate alternating currents. The alternating currents produced by a vacuum-tube oscillator are sometimes referred to as **oscillations,** and a tube that is operating as an a-c generator is said to be **oscillating.**

## *THEORY OF THE VACUUM-TUBE OSCILLATOR*

**Conditions Necessary to Sustain Oscillations.** Any ordinary amplifier circuit can be rearranged in a manner that will cause the amplifier tube to oscillate. As a matter of fact, a vacuum tube is made to oscillate solely by virtue of its amplifying properties.

Figure 122 is a diagram of an elementary single-tube amplifying stage. The input and output circuits are inductively (transformer) coupled to the grid and plate circuits, respectively, of the tube. If a sinusoidal alternating current is impressed on the input circuit and the tube is operating class A, the amplified output delivered to the load circuit will also be sinusoidal, or essentially identical in wave form to the input.

If 1 w of a-c power is supplied to the input circuit and 10 w of a-c power is delivered to the load, it can be said that 9 w of power has been supplied, or "generated," by the tube. This is not a violation of the law of conservation of energy one is not "getting something for nothing." The additional 9 w of a-c power has actually been taken from the plate supply or B battery.

From this point of view, therefore, the vacuum tube may be considered not an amplifier but a converter of direct current to alternating current. It can be compared to a motor generator having a d-c motor and an a-c generator. In the latter case, it is necessary that the alternator be provided with some form of external excitation power for the fields. Similarly, the vacuum tube requires excitation in the form of the original 1 w of power supplied to its input circuit.

A power amplifier of the foregoing type is often referred to as an **externally excited,** or **separately excited,** oscillator. *Master oscillator-power-amplifier* transmitter systems are sometimes designated as externally excited oscillator systems. Modern practice, however, is to identify separately the components of such a system. The power amplifier is not considered as a true externally excited oscillator but is treated as a simple power amplifier. The master oscillator is designated according to the method of excitation. The term "externally excited oscillator" is reserved for crystal oscillators. Such oscillators are discussed in a later section of this chapter.

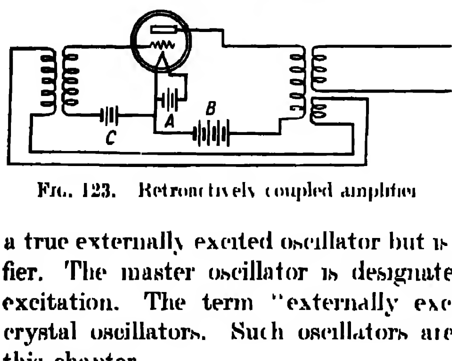


FIG. 123. Retroactively coupled amplifier.

In the amplifier circuit of Fig. 122 only 1 w of power is required to excite the tube, and an output of 10 w is obtained. If this circuit is modified as shown in Fig. 123, 1 w of the 10 w of output power can be fed back into the input circuit, thereby eliminating the need for external excitation power. The tube then becomes a true oscillator or generator, of alternating currents. Such an oscillator is called a **self-excited oscillator.**

The process of feeding power back into the grid circuit is called **feedback.** Feedback circuits are of three major kinds in common with other types of coupled circuits, namely inductively coupled feedback, conductively coupled feedback, and capacitively coupled feedback circuits. Figure 123 is an example of an elementary inductively coupled feedback circuit. In conductively coupled feedback circuits, the energy is conducted back to the grid circuit. In such a circuit, a single inductance usually serves both in the grid and the plate circuits. In capacitively coupled feedback circuits, energy from the plate circuit is fed back to the grid circuit through a capacitor, and in many types of oscillator

circuit, the interelectrode capacitance of the tube itself is utilized for this purpose. In general, coupling circuits that are utilized to feed back energy from the plate to the grid circuit of a tube are called **retroactively coupled circuits** or **regenerative circuits.**

In a power-amplifier circuit, the voltage output is 180° out of phase with the voltage input. In general, therefore, one of the conditions necessary to sustain self excited oscillations is that the power fed back into the grid circuit must be 180° out of phase with the output power. The frequency at which oscillations occur depends upon the inductance and capacitance constants of the plate circuit. In some oscillator circuits, the grid circuit is also composed of a tuned circuit of $L$ and $C$. In such oscillators in order that the voltage fed back to the grid circuit may be of the proper phase and amplitude to enable the tube to supply its own input, the plate circuit must be tuned to a frequency slightly higher than the frequency of the grid circuit. This frequency relation then is another of the conditions necessary to sustain oscillations in self-excited oscillators of this type. As a matter of fact, oscillations are often controlled in such circuits by detuning either the plate or grid circuit.

When, as is usually the case, it is desired to obtain appreciable power output from an oscillator, the tube is adjusted so that it operates as a class C amplifier. The output of such an oscillator is very little less than the corresponding class C amplifier. It differs by the amount necessary to supply driving power to the tube The main difference between a class C amplifier and an oscillator, aside from the circuit arrangement, is that the latter must utilize grid leak bias. Grid leak bias is necessary to make the oscillator self-starting and to ensure circuit stability.

If feedback is introduced in a simple amplifier circuit and the amount of energy fed back to the grid circuit is gradually increased from zero, it was seen above that a point is reached at which the grid-circuit losses are overcome and the tube oscillates. Between the moment of zero feedback and the point at which oscillations commence, the tube is said to be **regenerating.**

Regeneration has several desirable and several harmful effects. It is often purposely introduced into r-f amplifier circuits. Since a portion of the plate-circuit signal voltage is fed back into the grid circuit and reamplified several times, the over-all gain of an amplifier can be considerably increased by regeneration. In addition, since the feedback energy relieves the signal voltage of a portion of its function of supplying excitation to the stage, the effect is the same as a decrease in resistance of the input circuit. Under these conditions, the effective resistance is said to be decreased, but the effective $Q$ (see Chap. XII) is raised and results in a considerable increase in the selectivity of the circuit.

One of the serious disadvantages of regeneration in amplifiers is that the increased selectivity tends to cause suppression of the higher

side-band frequencies of the signal with resultant distortion. Furthermore, regenerative amplifiers are notoriously unstable, require adjustment whenever the circuit frequency is changed, and are likely to break into oscillation.

Regeneration finds its greatest application in detector circuits designed for the reception of radiotelegraph signals. In such circuits, the regenerative detector affords considerable amplification for the interception of modulated carrier telegraph signals. At the same time by increasing the feedback, the detector is made to oscillate, thus affording heterodyne oscillations for the reception of c-w telegraph signals. This condition is discussed further in the chapter on receiving circuit principles.

**Harmonics.** Alternating currents that are perfectly sinusoidal in wave form are practically never encountered in actual radio circuits. Although vacuum-tube oscillators are capable of generating alternating currents that are *almost* perfectly sinusoidal, there is always *some* distortion of the wave form. In the usual case of such distortion, the generated wave contains components of two, three, or more frequencies, all of which are integrally related to the original frequency.

The frequency of the original generated wave is called the **fundamental frequency,** but it is also often referred to as the **first harmonic,** since a pure sinusoidal wave is a form of simple harmonic motion. The additional component frequencies are always multiples of the fundamental. Thus, an oscillating vacuum tube, in addition to the fundamental frequency, may generate currents of *twice* the fundamental frequency, called **second harmonics,** currents of *three* times the fundamental frequency, called **third harmonics,** and so on. Some vacuum-tube oscillator circuits are especially rich in harmonics and are used for laboratory instruments in which the harmonic output is utilized. The multivibrator circuit discussed later in this chapter is an example of an especially prolific harmonic generator. Harmonics as high as the eightieth can be generated with this oscillator.

The lower-order harmonics, second, third, and so on, usually have the greatest amplitudes. The succeeding harmonics fall off quite sharply in amplitude after the third, and the higher order harmonics become progressively weaker.

The effective value of amplitude (either current or voltage) for a complex wave, containing several harmonics in addition to the fundamental, will vary considerably from the effective value of a pure sinusoidal wave. Specifically, the effective value of a complex wave of this nature is equal to the square root of the sum of the squares of the effective values of the individual component frequencies. This relation holds true whether or not the components of a complex wave are harmonically related. Expressed mathematically,

$$E_{\text{eff}} = \sqrt{E_1^2 + E_2^2 + E_3^2 + \cdots}, \tag{1}$$

where $E_{eff}$ = effective value of complex wave;
$E_1$ = effective value of fundamental;
$E_2$ = effective value of second harmonic;
$E_3$ = effective value of third harmonic,
and so on.

There are many instances where the harmonics generated by a vacuum-tube oscillator are put to great practical use in radio circuits. More often, however, in both receiving and transmitting circuits, harmonics are undesirable, and special circuit design is necessary to eliminate them. Both the utilization and the suppression of harmonics are discussed as the particular application is encountered in this text.

**Heterodynes.** One of the most useful phenomena occurring in a-c circuits is that of heterodyning. In an a-c circuit, when two alternating voltages of different frequencies are impressed upon the same circuit in such a way that the output contains the product of the two voltages, additional frequencies, called **heterodyne frequencies,** or, simply, **heterodynes,** result *Four* frequencies then exist in the circuit: the two original frequencies and *two* heterodyne frequencies. One heterodyne will have a frequency equal to the *sum* of the two original frequencies. The other heterodyne will have a frequency equal to the *difference* of the two original frequencies.

Thus, if two alternating currents of 9,000 and 9,500 c are impressed upon a circuit, two additional frequencies will be created in the circuit. One heterodyne frequency will be 18,500 c, or the sum of the original frequencies. The other heterodyne will be 500 c or the difference of the two original frequencies. *Four* different frequencies will therefore be present in the circuit: 9,000 c; 9,500 c 18,500 c, and 500 c. The difference frequency is customarily called the **beat frequency.** The process of combining two such currents of different frequencies to produce additional frequencies is called **heterodyning,** or **beating,** the two frequencies together.

The principle of changing frequency by means of heterodyning has a number of important applications in radio work. It is the fundamental principle upon which the superheterodyne receiver (discussed in Chap. XII) operates. Heterodyning is also used in telegraph receivers to make an r-f signal audible to human ears. Thus, if a 1,000 kc signal is impressed upon a sound-conversion device, such as an earphone, the sound waves set up will vibrate at a frequency of 1 000,000 times per second, far beyond the range of the human ear. If, however, the 1,000-kc signal is beat with a 1,001-kc current generated locally and the circuit is connected to an earphone, the difference frequency of 1,000 c will set up 1,000-c sound waves, which, of course, are easily heard by the human ear. If the incoming 1,000 kc signal is interrupted in the dots and dashes of the Morse code, the beat frequency will also be interrupted in like

manner, since it depends for its existence upon *both* of the original frequencies.

## *STANDARD OSCILLATOR CIRCUITS*

A number of standard oscillator circuits have been developed for the production of oscillations. In general, these circuits differ mainly in the methods of coupling to the load and in the methods of retroactive coupling. Special circuits have also been devised that have advantages peculiar to certain applications.

Any standard oscillator circuit may be classified according to the method of feeding the d-c power to the circuit as a series feed circuit or a shunt feed circuit. These methods should not be confused with feedback. Series feed circuits are those in which the plate supply voltage,

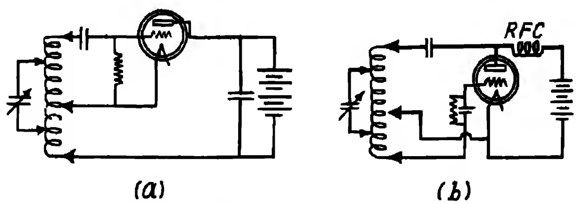



FIG. 124. The Hartley oscillator (a) Series feed. (b) Shunt feed.

the tube, and the plate-circuit parallel resonant circuit (customarily called the **tank circuit**) are all connected in series. Shunt feed circuits are those in which the plate supply voltage, the tube, and the tank circuit are connected in parallel.

In series feed circuits, it is necessary to by-pass the plate-supply source with a suitable capacitor in order to prevent the a-c plate-circuit component from flowing through it. The tank circuit is also at a high d-c potential with respect to the filament, which is undesirable.

In shunt feed circuits, an r-f choke coil is required between the power supply and the tube circuit and effectively excludes the a-c component from the power source. A blocking capacitor is utilized between the tube and the tank circuit to exclude high d-c potentials from the tank circuit, to prevent the tank-circuit inductance from short-circuiting the power supply, and to afford a path of low reactance to the desired a-c component. Shunt-feed oscillators are usually preferred to series feed for most applications.

**The Hartley Oscillator.** The Hartley oscillator circuit is very widely used mainly because a single inductance serves for both grid and plate circuits. Diagrams of both series- and shunt-feed Hartley oscillator circuits are shown in Fig. 124. An intermediate tap on the inductance is connected to the filament and divides the inductance effectively into a grid coil and a plate coil.

In this circuit, a portion of the voltage developed in the plate coil is fed back to the grid circuit both by inductive and conductive coupling. The series-feed Hartley is not recommended for other than low-power battery-operated oscillators because of the shunting effect of any plate to filament capacitance inherent in the power supply. Such capacitance is usually found between the windings of filament transformers and the plate winding of the power transformer. Since this capacitance is applied directly across the plate coil, it directly affects the circuit tuning. In addition, undesirable high r-f voltages will be impressed across the filament-plate windings of the power-supply transformers.

The Hartley circuit is usually inductively coupled to the load circuit at the plate end of the inductance.

**The Armstrong Oscillator.** Diagrams of both series- and shunt-feed

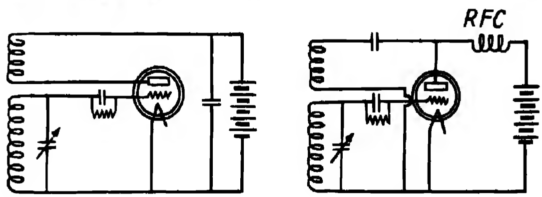



FIG. 125. The Armstrong oscillator. (a) Series feed (b) Shunt feed.

Armstrong oscillators are shown in Fig. 125. The Armstrong oscillator differs from other types mainly in that the grid circuit is tuned instead of the plate circuit

Feedback is obtained by inductive coupling between the untuned plate circuit and the tuned grid circuit. Usually this coupling is made variable, permitting adjustment for maximum stability and output. The load circuit can be inductively coupled to the grid coil. In the series feed Armstrong, the output circuit is often coupled directly to the plate circuit by inserting a primary inductance in series with the plate inductance. The plate inductance in an Armstrong oscillator is usually called a **tickler coil,** and for this reason, this circuit is often referred to as the **tickler feedback circuit.**

The Armstrong oscillator finds its greatest application in oscillating detector circuits.

**The Tuned-plate-tuned-grid Oscillator.** In this type of oscillator, both the grid and plate circuits contain parallel resonant $L$ and $C$ circuits. When the plate circuit is tuned to a frequency slightly higher than that of the grid circuit, the tube oscillates: and the plate-circuit energy is then fed back to the grid circuit through the interelectrode capacitance of the tube. Diagrams of series- and shunt-feed tuned-plate tuned-grid oscillators are shown in Fig. 126. This type of oscillator is very popular for low-power transmitters because of the relative ease of adjustment.

The load circuit is customarily coupled to the plate coil of the tuned-plate–tuned-grid oscillator.

**The Colpitts Oscillator.** The Colpitts oscillator is similar to the Hartley circuit. Energy from the plate circuit is fed back to the grid circuit by capacitive coupling, however, instead of by inductive coupling, as in the Hartley. The tank-circuit capacitance is made up of two

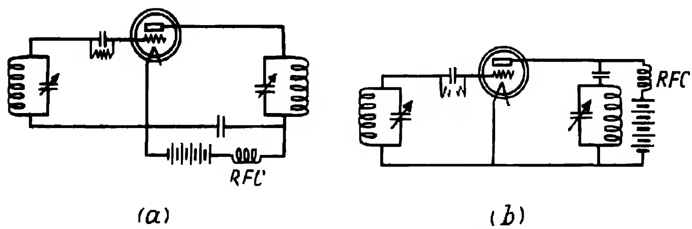



Fig. 126. The tuned-plate tuned-grid oscillator. (a) Series feed. (b) Shunt feed.

capacitors in series with their common connection going to the filament. The excitation voltage fed to the grid is adjusted by changing the value of the grid capacitor. Since this capacitor is part of the oscillatory circuit, any change in its capacitance will also affect the frequency of oscillations. Excitation adjustments must, therefore, be compensated for by a change in the plate capacitor.

Although not as popular as the Hartley oscillator, the Colpitts circuit is widely used because of its adaptability. With the proper feedback

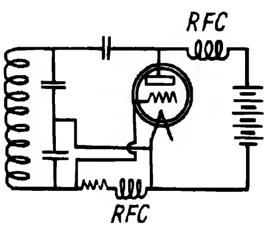



Fig. 127. The Colpitts oscillator.

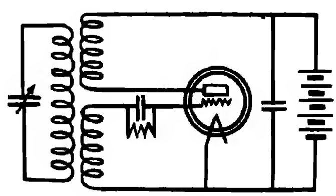


Fig. 128. The Meissner oscillator.

ratio established by the use of fixed capacitors, a wide frequency range can be covered by utilizing a variable inductance without disturbing the feedback condition. As in the Hartley oscillator, the load circuit is coupled to the oscillator coil inductively. A circuit diagram of a shunt-feed type of Colpitts oscillator is shown in Fig. 127.

**The Meissner Oscillator.** The Meissner oscillator differs from other types in the use of an additional inductance coupled to both grid and plate inductances of the tube. This mutual inductance with its shunt condenser forms the actual oscillatory circuit. A diagram of a Meissner oscillator circuit is shown in Fig. 128.

By judicious adjustment of the coupling and variation of the tank-circuit constants, the Meissner oscillator can be made to operate more efficiently than many other types of self-excited oscillator. Nevertheless, because of the number of coils necessary and the difficulty of proper adjustment, the Meissner oscillator is not so widely used as other types.

**The Electron-Coupled Oscillator.** In all the foregoing types of self-excited oscillators coupling a load to the oscillator definitely affects circuit conditions. Coupling a load to the inductance of an oscillator tends to decrease the effective impedance of the circuit. Any variation in tank circuit impedance tends to disturb the voltage-feedback ratio and also affects the frequency of oscillations. When adjusting a coupled oscillator, any change in load coupling or load value must therefore be accompanied by a readjustment of the oscillator circuit proper in order to maintain peak efficiency.

In earlier types of transmitters, oscillators delivering the required power were coupled directly to the antenna system. Such transmitters were notoriously unstable because of the variable load reflected by the antenna. Any movement of the antenna (swinging in the wind, and so on), dampness, variation in antenna losses due to weather conditions, and so on, changed the effective impedance and thus reflected a varying load into the oscillator circuit. The result was a variation in the frequency generated by the oscillator tube.

To overcome this difficulty, the *master-oscillator power-amplifier* transmitting circuit was developed. This circuit (commonly abbreviated MOPA) consists simply of an oscillator coupled to a power amplifier. In modern transmitters a number of power amplifiers are usually employed. Such transmitting circuits will be discussed separately in Chap. XIII.

The final power amplifier in an MOPA system is coupled to the antenna or radiating system. Any change in antenna load conditions does *not* affect the frequency of the signal transferred from the power amplifier to the antenna, since the frequency depends upon oscillator-circuit conditions alone. The power amplifier input circuit presents an essentially constant load to the oscillator circuit regardless of antenna-load conditions. The oscillator frequency is therefore independent of antenna-load conditions.

The introduction of the screen-grid tube presented an opportunity for the development of a master-oscillator power-amplifier circuit utilizing a single tube. In this circuit, the control and screen grids of the tube are used as the grid and plate, respectively, for a triode oscillator. The variations in the electron stream caused by the generation of oscillations in this circuit produce an a-c component in the real plate circuit of the screen-grid tube. The frequency of this a-c component is a direct function of the oscillator frequency. The plate circuit of the tube is,

therefore, said to be **electronically coupled** to the oscillator circuit. The load can be coupled capacitively or inductively to the plate circuit. With this system, the frequency of oscillations is practically independent of load variations. This circuit was originally designed by B. J. Dow in 1931 and is called the **electron-coupled oscillator circuit.** A diagram for this popular type of oscillator is shown in Fig. 129.

The main advantage of the electron-coupled oscillator is its frequency stability. It is not readily adaptable, however, to circuits utilizing high-power tubes, and for this reason, its use in transmitting circuits is comparatively limited

Moreover, since the master-oscillator power-amplifier transmitting circuit is in practically universal use, the need for electron-coupled oscillators is obviated. Frequency stability under varying load conditions is established by the successive power amplifier stages, regardless of the type of oscillator used.

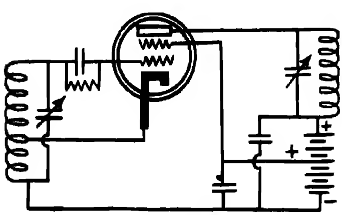


FIG. 129. The electron coupled oscillator.

The electron coupled oscillator, therefore, finds its greatest application in receiving circuits, such as the superheterodyne (see Chap. XII). In such circuits since the oscillator stage contributes nothing toward the gain of the receiver, space limitation is an important factor Many circuits incorporate the functions of first detector and oscillator within a single tube acting as an electron coupled oscillator.

The electron coupled oscillator is an ideal solution for receiver applications. The single tube reduces the space requirements, and, by proper choice of the tube and operating voltages the power output can be adjusted to any desirable value

It should not be inferred that utilizing a power-amplifier stage in a transmitter automatically provides frequency stability. Such an arrangement *does* prevent varying antenna load conditions from affecting the oscillator frequency. However, the oscillator frequency is still subject to variation due to inherent faults in the oscillator circuit itself.

In general, a number of factors contribute to frequency variation in self-excited oscillator circuits. Frequency instability can be caused by variation of inductance and capacitance by temperature, variations in the interelectrode capacitance of the tube (due mostly to changes in tube temperature), and by variations of the reflected load impedance caused by any of the above factors affecting a succeeding amplifier stage. In addition, any variation in plate supply and bias and filament voltages can cause a change in the frequency of the oscillations generated by the tube.

**The Crystal Oscillator.** Stability in h-f oscillators has been obtained by the use of the quartz-crystal oscillator circuit. This type of oscillator depends for its operation upon the so-called piezoelectric properties of certain crystalline substances, including Rochelle salts, tourmaline, quartz, and others. Because of its mechanical ruggedness, low temperature coefficient, and cheapness compared with the other piezoelectric crystallines, quartz is used exclusively in oscillator circuits.

A natural quartz crystal is shown in Fig. 130. A complete quartz crystal is very rarely found. Such a crystal would have both ends pointed and have a hexagonal cross section. The properties of such a crystal are customarily expressed in terms of three sets of axes. The axis joining the points of the crystal at the ends is called the **optical axis.** The three axes $X_1$, $X_2$, and $X_3$ (see Fig. 131(a) ) are at right angles to the optical axis and join the corners of the hexagon which forms the cross section of the crystal These are called the **electrical axes.** The three axes $Y_1$, $Y_2$, and $Y_3$, shown in Fig. 131(b), join opposite sides of the hexagonal cross section and are at right angles to the optical axis Each $Y$ axis is also at right angles to one of the $X$ axes The $Y$ axes are called the **mechanical axes** of the crystal.

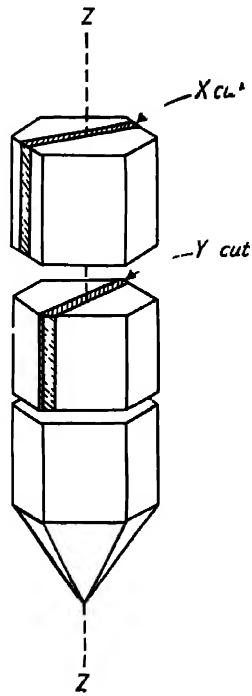


FIG. 130 A natural quartz crystal

If a flat section is cut from a crystal in such a way that the flat sides are perpendicular to an electrical or $X$, axis, as shown in Fig. 131(b), this section of crystal is found to have peculiar properties If a mechanical stress is applied to the crystal along a $Y$ or mechanical, axis, it will produce electric charges on the flat sides of the crystal section. If the direction of the mechanical stress is changed from compression to tension, or vice versa, it causes a reversal of the polarity of the charges on the flat sides of the crystal. Thus, if mechanical vibrations, that is, alternate compressions and tensions, are applied to the $Y$ axis of the crystal, an alternating emf is produced across the flat faces, or along the $X$ axis, of the crystal. The frequency of the alternating emf so produced will be directly dependent upon the frequency of the mechanical vibrations applied.

Conversely, if a difference of potential is applied across the faces (along the $X$ axis) of the crystal, a mechanical stress along the $Y$ axis will be produced. This stress will result in either compression or expansion of

the crystal along this axis depending upon the polarity of the applied voltage. Thus, if a given voltage results in an expansion of the crystal, reversing the polarity will result in a compression of the crystal. Hence, it follows that if an alternating voltage is applied across the faces of the crystal, a mechanical vibration of the crystal will be set up because of the successive alternate compression and expansion along the $Y$ axis.

A piece of quartz, like every other rigid object, has a natural period of vibration. When an alternating voltage is applied to a crystal, as above, the crystal will vibrate, although the amplitude of vibrations will be comparatively small. When the frequency of the impressed alternating emf coincides with the natural mechanical period of vibration of the

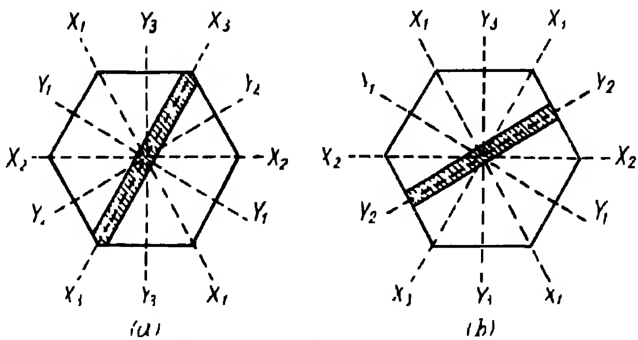



FIG. 131. Quartz crystal sections showing $Y$ and $X$ cuts. (a) $Y$ cut (30° cut) crystal section. (b) $X$ cut (Curie cut) crystal section.

crystal, the amplitude of the vibrations is very greatly increased. If the voltage impressed on the crystal is large enough, the crystal can be made to shatter itself.

The peculiar properties of quartz outlined above are called **piezoelectric** properties. The Curies were the first to utilize quartz crystals for their piezoelectric qualities. They used $X$ cut crystals for the calibration of electrometers. For this reason, $X$ cut crystals are often called "Curie-cut" crystals.

The application of quartz crystals as frequency stabilizers in oscillator circuits was developed by W. G. Cady in 1922. He showed how a quartz crystal could be treated as a form of electric series resonant circuit. Thus, if a thin slab of quartz crystal has metallic plates containing its opposite faces, it would appear, offhand, to have the characteristics of an efficient capacitor, the two metallic plates forming the plates of a capacitor and the quartz, having an extremely high insulation resistance, forming an efficient dielectric. If connected to an a-c bridge, it would be expected that the quartz would show a constant capacity with negligible series resistance. The actual result of connecting a quartz crystal so

arranged to an a-c bridge, however, is strikingly different. Actually, the crystal exhibits properties very similar to a series resonant circuit. As the frequency of the impressed alternating emf is increased, the crystal displays a changing capacitive reactance. At the frequency that coincides with the natural mechanical period of the crystal, the impedance becomes a minimum and becomes a pure resistance Within a certain narrow band of other frequencies, the crystal actually shows inductive reactance. The variation of reactance with frequency is shown graphically in Fig. 132.

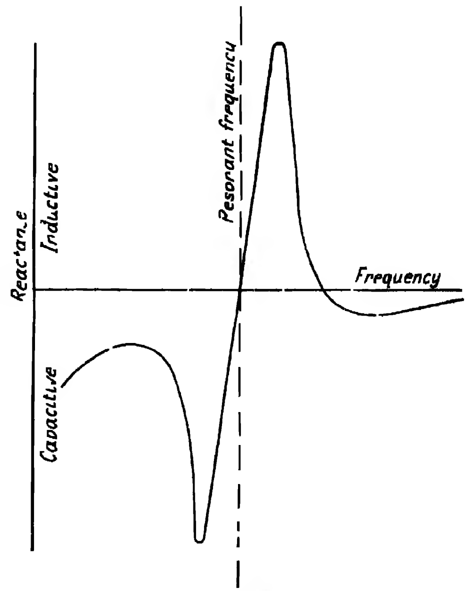



FIG. 132 Electrical characteristics of a quartz-crystal plate as the impressed electrical frequency passes through crystal resonance.

As the quartz crystal is utilized in oscillator circuits, it takes the place of the tuned resonant circuit in the tuned plate tuned grid oscillator circuit (Fig. 126). The crystal is placed in a holder which consists, essentially, of two metallic plates parallel to the opposite flat faces of the crystal. A small air gap is provided between one plate and the crystal face to allow it to vibrate freely.

Although the crystal itself is equivalent to a series resonant circuit, the combination of the crystal and vacuum tube may be considered as a parallel resonant circuit, in which the grid capacitance of the tube resonates with the apparent inductance of the crystal. Since the crystal presents an inductive reactance only within an extremely narrow range of frequencies, the circuit can operate as a tuned-plate-tuned grid oscillator only within this range

Thus, the quartz crystal has the effect of a very sharply resonant circuit. Since it is not subject to the many factors that contribute to frequency variation of an ordinary tuned resonant circuit, the quartz crystal has a great stabilizing effect on the frequency of oscillations produced.

The resonant frequency of a crystal is dependent almost completely upon its physical dimensions. Anything, therefore, which does *not* affect the dimensions of the crystal can have very little effect on the frequency. Thus, variations of circuit voltages, load conditions, and so on, although

they may affect the *amplitude* of oscillations, will not affect their frequency to an appreciable extent.

Strictly speaking, there are several factors which may affect the resonant frequency of a quartz crystal, such as temperature, plate capacity, vibration. The one factor that produces a frequency shift large enough to become troublesome is the temperature. Crystals in commercial transmitters are accordingly mounted in ovens having the temperature automatically controlled, usually by thermostatic means.

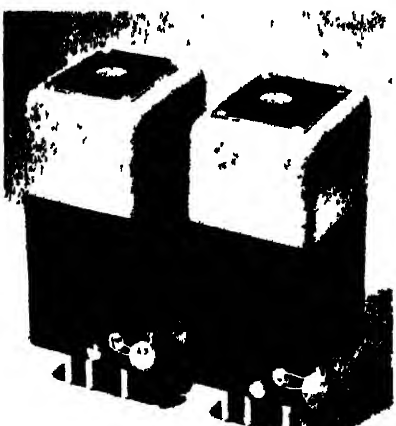


FIG. 133. Duplicate crystal ovens—broadcast transmitter. (*Courtesy of RCA Inc.*)

*X*-cut crystals have a negative temperature coefficient, causing the natural frequency to decrease as the temperature rises. The decrease amounts to about 10 to 25 parts in a million per degree centigrade. The *Y*-cut crystals have a positive temperature coefficient, the natural frequency rising with increase in temperature. The increase in frequency varies from about 25 to 100 parts in a million per degree centigrade.

If the plane of a crystal section is rotated about the *X* axis to make an angle of approximately 35° with the optical axis, a new cut of crystal can be obtained, which is called the **AT cut.** The *AT*-cut crystal combines the characteristics of the *X*-cut and *Y*-cut crystals. As a result, its resonant frequency is substantially independent of temperature changes. *AT*-cut crystals are in wide use for frequencies above 300 kc.

Many different circuit arrangements can be employed to utilize a quartz crystal as the frequency controlling element in a vacuum tube oscillator. The simplest of these is the triode circuit shown in Fig. 134.

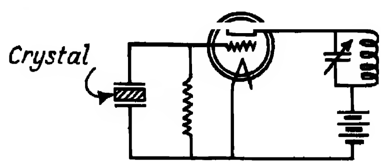



FIG. 134. Triode crystal oscillator.

This circuit is the equivalent of the tuned-plate tuned-grid circuit. The resonant circuit properties of the crystal are used to replace the tuned-grid tank circuit. The two plates of the crystal holder are connected between the grid filament terminals of a small triode power tube having a parallel resonant circuit in the plate circuit. When the plate tank circuit is adjusted to approximate resonance, the feedback through the grid-plate

interelectrode capacitance of the tube excites the grid circuit and oscillations begin. The frequency is dependent upon the mechanical period of vibration of the crystal. Consequently, the plate circuit may be considerably detuned without materially affecting the frequency.

A negative bias is sometimes necessary in crystal-oscillator circuits and can be applied through the grid leak. In general, small power oscillators operating at moderate values of plate voltage will operate without bias.

The power obtainable from a crystal oscillator is limited by the power-dissipation capabilities of the crystal and by the ease with which the crystal will shatter. The latter factor is dependent upon the thickness of the crystal. In general, the power output is limited at high frequencies by heating of the crystal and at low frequencies by the mechanical stresses set up by the vibrations that may result in cracking the crystal.

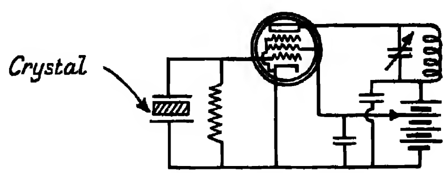


FIG. 135 Pentode crystal oscillator.

Since the power output is limited by the amount of power that can be fed back to the crystal, it follows that the greatest power output can be obtained without danger to the crystal by utilizing a tube of high power sensitivity. The power pentode or beam tetrode tubes find wide application as crystal oscillators for this reason. In addition, the screen grid in these tubes decreases the grid plate capacitance, thus limiting the amount of feedback energy and permitting the use of higher plate voltages without overloading the crystal. A circuit of a typical pentode-type crystal oscillator is shown in Fig. 135.

The circuit most commonly used in commercial equipment is the triode crystal oscillator of Fig. 134. A typical commercial transmitter usually has a number of power amplifier stages following the oscillator stage. It is seldom necessary to develop much power in the crystal stage. Oscillator power output is therefore usually sacrificed in the interest of circuit stability.

## SPECIAL OSCILLATOR CIRCUITS

**The Dynatron Oscillator.** It was shown in Chap. X that when the grid immediately adjacent to the plate of a vacuum tube is made more positive than the plate, excessive secondary emission from the plate occurs. If the surface condition of the plate and the plate voltage are such that this emission is appreciably large, the tube plate-filament resistance becomes effectively negative. A tube operated under these conditions is called a **dynatron.**

The dynatron is often utilized as an oscillator. If a parallel resonant circuit is inserted in the plate circuit of the tube, oscillations will start when the absolute value of the negative resistance is less than the parallel resonant impedance of the plate tank circuit. Thus if the circuit voltages are properly adjusted, the number of secondary electrons leaving the plate exceeds the number of primary electrons arriving at the plate, and

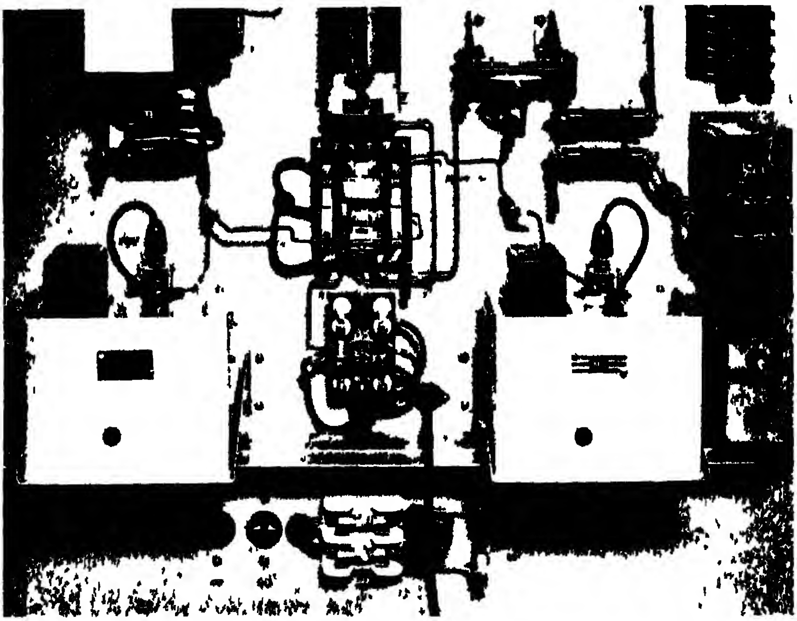


Fig. 136 Duplicate crystal oscillator units in modern broadcast transmitter (*Courtesy of RCA Inc.*)

the plate current reverses, despite the positive plate potential. The circuit is shown in Fig. 137.

A screen-grid tube as shown in the illustration, is usually more satisfactory as a dynatron oscillator because of the greater frequency stability.

Dynatron oscillators are used mainly for laboratory or test oscillators. They are especially adaptable as radio-frequency oscillators. Remarkably high frequency stability with variation of tube voltages has been obtained on frequencies up to 30 megacycles.

**The Multivibrator.** Another popular type of laboratory oscillator is the **multivibrator,** sometimes called the **relaxation oscillator.** A typical multivibrator circuit is shown in Fig. 138. It consists of a two stage resistance-coupled amplifier circuit in which the output voltage of the second tube is fed back into the grid circuit of the first tube. Since each tube produces a phase shift of 180°, the output of the second tube supplies

an input voltage to the first tube that has the right phase to sustain oscillations.

Since there is no tuned circuit, the frequency of oscillation is governed primarily by the grid-circuit resistance and capacitance. It is also affected by other circuit constants, tube voltages, and tube characteristics. Frequencies as low as 1 c per *minute* and as high as 100,000 c per second can be generated by a multivibrator.

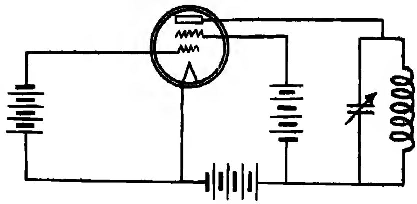


FIG. 137 The dynatron oscillator

Multivibrators are widely used in frequency-measuring equipment because of the rich harmonic-output content. Harmonics as high as the eightieth can be detected without the use of external amplifiers. Another advantage is the wide frequency range that can be covered by means of a single adjustment.

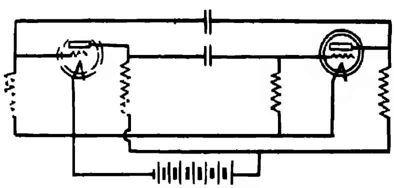


FIG. 138 The multivibrator

**The Magnetostriction Oscillator.** The magnetostriction oscillator is another type of laboratory oscillator used to generate low frequencies where frequency stability is important. The circuit is shown in Fig. 139 The circuit depends for its operation upon the natural mechanical period of vibration of a bar of iron alloy. Nickel, alloys of nickel and iron, Invar, and Nichrome have pronounced magnetostrictive effects.

It is seen that the circuit is similar to the Hartley oscillator except that the mutual inductance between the coils is reversed. The magnetostrictive rod is clamped at the center and is magnetized by the d-c component of the tube plate current. Since the coils are reversed, the circuit will not oscillate without the rod: but when the rod is inserted, it provides the coupling between the coils, and oscillations start. The rod vibrates longitudinally at a frequency dependent upon the physical characteristics of the rod. The tank circuit capacitor can be varied over a considerable range without affecting the oscillator frequency.

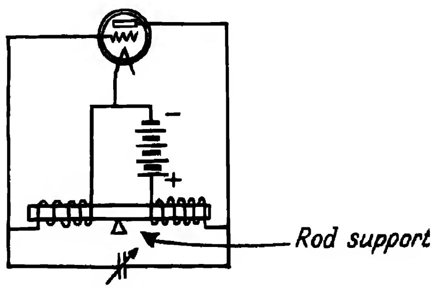



FIG. 139 The magnetostriction oscillator.

**Parasitic Oscillations.** Any *undesired* oscillation occurring in an oscillator or power amplifier is called a **parasitic oscillation.** Such

unwanted oscillations result from the fact that several different modes of oscillation can exist in a circuit in addition to the desired types of oscillation. These modes are due to parasitic circuits formed by tube capacitances, stray inductance, and capacitance of lead wires, chokes, and so on. Since such oscillations absorb power, they represent a distinct loss in the circuit and must be avoided. In addition, the voltages set up in a circuit by such oscillations introduce distortion and other undesirable effects.

Parasitic oscillations can be minimized by using simple circuits and by making interconnecting wires as short as possible. Screen-grid, or pentode, oscillators are less conducive to parasitics if the screen grid is operated at ground r-f potential, and they are often preferred to triode oscillators for this reason. Parasitics are also less prevalent in circuits utilizing inductive output coupling than in circuits using other coupling methods.

**U-H-F Oscillators.** The production of oscillations at extremely high frequencies presents several problems that require special circuit design. At such frequencies, the period of oscillations becomes comparable to the time of transit of the electrons within the tube. As a result, pronounced difficulties occur in attempting to secure the proper phase relations between the alternating grid and plate voltages.

## *QUESTIONS AND PROBLEMS**

**1.** What are the advantages of a master-oscillator power-amplifier type of transmitter as compared to a simple oscillator transmitter?

**2.** Draw a simple schematic diagram of a pentode-type tube used as a crystal-controlled oscillator, indicating power-supply polarities.

**3.** What are the differences between Colpitts and Hartley oscillators?

**4.** Draw a simple schematic diagram showing a Colpitts-type triode oscillator with shunt-fed plate. Indicate power-supply polarity.

**5.** By what means is feedback coupling obtained in a tuned-grid-tuned-plate type of oscillator?

**6.** Draw a simple schematic diagram of a dynatron type of oscillator, indicating the circuit elements necessary to identify this form of oscillatory circuit.

**7.** What may be the result of parasitic oscillations?

**8.** What are the principal advantages of crystal-control oscillators over tuned-circuit oscillators?

**9.** Describe a multivibrator and list its characteristics and uses.

**10.** List the characteristics of a dynatron type of oscillator.

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

## Chapter XII

# RECEIVING-CIRCUIT PRINCIPLES

A modern radio-receiving system consists primarily of a system of vacuum tubes designed to *separate* the desired a-f component of a received radio signal from the nonaudible r-f component, to *amplify* the relatively feeble signal voltage to usable proportions, and to *convert* the final amplified a-f electric energy to energy in the form of sound waves to which the human ear is sensitive. The process of separating the a-f and r-f components of the signal is called **detection,** or **rectification.** The process of amplification is sometimes accomplished *prior* to detection, sometimes *after* detection, and often both before and after detection. The process of converting a-f electric energy to energy in the form of sound waves is the final step in a receiving system. Chapter XIV is entirely devoted to the theory of this sound conversion. When amplification is accomplished *before* detection, it is done by means of an *r-f amplifier.* When accomplished *after* detection, it is done by means of an *a-f amplifier.* All the various processes outlined above are combined to make up the complete unit called the receiver. In modern receivers, each of these processes is accomplished by means of one or more vacuum tubes. Among other things, a vacuum tube requires for its proper operation a source of direct current to be applied both to the plate to supply a constant positive potential and also to the grid as bias. Early receivers utilized batteries for this purpose, but modern receivers almost universally utilize *power packs.* Power packs function to rectify and to filter (smooth out) alternating current and make it suitable for use in vacuum tube circuits. Power packs are the outgrowth of the universal availability of alternating current. Since power packs are an integral part of practically all receivers, it is fitting that the principle of such power units be discussed first and the remaining receiver processes be taken in their order.

### *POWER PACKS*

**Primary Sources of Voltage.** The source of alternating current usually available is of comparatively low voltage. Thus, the 110-v a-c power available in the average home is too low for direct application to all but comparatively small radio receivers. In order to supply the higher voltages needed to operate most vacuum tubes, this primary voltage must be increased, or stepped up, by means of a transformer. The

secondary output of the transformer is then rectified, filtered, and applied to the receiving circuits.

In large receivers, this function is performed by a separate transformer called the **plate transformer,** since it ultimately supplies power to the plates of vacuum tubes.

In addition to the power necessary for plate consumption, it is also necessary to provide a means of heating the filaments of the receiver vacuum tubes. In a-c-operated receivers utilizing indirect heater-type cathodes, alternating current may be applied directly to the filaments, as described in Chap. X. The usual filament construction embodies filaments of low voltage consumption and comparatively high current consumption, the commonly employed heater voltages being 2.5, 5, and 6.3 v. Such low voltages necessitate the use of a step-down transformer to decrease the line voltage of 110 v to the proper value. Such a transformer is customarily referred to as a **filament transformer.**

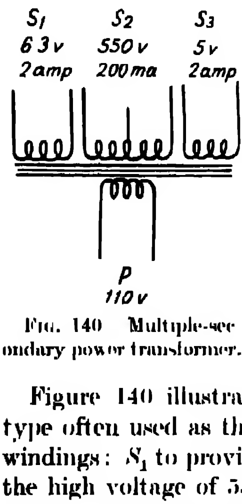


FIG. 140 Multiple-secondary power transformer.

Most modern receivers employ a single transformer with multiple secondary windings combining the functions of plate and heater-supply transformers. The theory of transformers has been discussed in Chap. VII. Multiple secondary transformers are simply extended applications of single secondary transformers. The same relation of turns ratio to voltage ratio exists, but the current ratio is determined by the load on the individual secondary windings.

Figure 140 illustrates a multiple secondary power transformer of a type often used as the basis for a receiver power pack. There are three windings: $S_1$ to provide 6.3 v for the receiver tube heaters; $S_2$ to provide the high voltage of 550 v for the plate supply, and $S_3$ to provide a 5 v source to heat the rectifier tube filament. The turns ratio of each of the various windings to the primary winding $P$ is directly dependent upon the voltage ratio in each case. Thus, the ratio of $S_1$ to $P$ is

$$\text{turns ratio} = \frac{S_1}{P} = \frac{6.3}{110} = 0.057. \qquad (1)$$

If the primary winding has 500 turns, therefore, the number of turns in secondary $S_1$ is found to be

$$\text{secondary turns } S_1 = 500 \cdot 0.057 = 28.5. \qquad (2)$$

The turns ratio of $S_2$ is

$$\text{turns ratio} = \frac{S_2}{P} = \frac{550}{110} = 5. \qquad (3)$$

For the same primary winding mentioned above (500 turns), secondary $S_2$ will then have

$$\text{secondary turns } S_2 = 500 \cdot 5 = 2{,}500 \text{ turns.} \tag{4}$$

Similarly, the turns ratio of $S_3$ is

$$\text{turns ratio} = \frac{S_3}{P} = \frac{5}{110} = 0.045. \tag{5}$$

For the same 500-turn primary winding the turns in secondary $S_3$ will be

$$\text{secondary turns } S_3 = 500 \cdot 0.045 = 22.5 \text{ turns.} \tag{6}$$

The total current drain of the primary is found by working backward from the load. Thus, if the current drain in secondary $S_1$ is 2 amp, in $S_2$ 200 ma, and $S_3$ 2 amp, the primary current is found as follows:

$$\left.\begin{aligned}\text{primary current due to } S_1 &= 2 \cdot 0.057 = 0.114 \text{ amp}\\ \text{primary current due to } S_2 &= 0.2 \cdot 5 = 1 \text{ amp}\\ \text{primary current due to } S_3 &= 2 \cdot 0.045 = 0.09 \text{ amp}\end{aligned}\right\}. \tag{7}$$

Therefore,

$$\text{total primary current} = 0.114 + 1 + 0.09 = 1.204 \text{ amp.} \tag{8}$$

A check can be made of the above calculation by use of the formula

$$P_p = P_1 + P_2 + P_3, \tag{9}$$

where $P_p$ = power input to primary;
$P_1$ = power output of secondary $S_1$,
$P_2$ = power output of secondary $S_2$,
$P_3$ = power output of secondary $S_3$.

Then,

$$\text{primary power} = 110 \cdot 1.204 = 132.4 \text{ w.} \tag{10}$$

$$\text{secondary } S_1 \text{ power} = 6.3 \cdot 2 = 12.6 \text{ w.} \tag{11}$$

$$\text{secondary } S_2 \text{ power} = 550 \cdot 0.2 = 110 \text{ w.} \tag{12}$$

$$\text{secondary } S_3 \text{ power} = 5 \cdot 2 = 10 \text{ w.} \tag{13}$$

Substituting the above power value in Eq. (9),

$$P_p = 12.6 + 110 + 10 = 132.6 \text{ w.} \tag{14}$$

Considering that the turns ratios were computed to only three decimal places, this value falls close enough to the previously calculated value of 132.4 w to provide a satisfactory check. It should be remembered that the preceding treatment has been on the assumption that no losses occur in the transformers. In practical applications the efficiency of the transformer must be taken into consideration.

Many of the smaller receivers in use at the present time embody power packs that utilize no transformer whatsoever. In some such cases, the

plate supply is obtained by rectifying directly the 110 v obtainable from the power lines. The heaters of the tubes are connected in series with a suitable resistor to maintain the proper voltage on each tube. Since the current is the same in all parts of a series circuit, the vacuum tubes used in such a receiver must have identical heater current ratings. A number of tubes have been developed having unusually high filament-voltage ratings, especially for this application. The tube complement of the receiver is then so chosen that no series resistor is necessary.

The above system has the serious disadvantage of the low plate voltage available. A scheme to overcome this disadvantage is the so-called voltage-doubler rectifier system. This is described in detail in the following section on the rectifier.

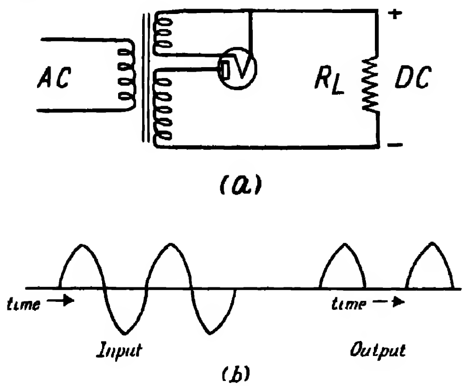



FIG. 141 (a) Elementary half wave rectifier circuit (b) Input and output wave forms for circuit of (a)

**The Rectifier.** Perhaps the most important unit in a power pack is the rectifier. The rectifier converts the alternating current obtained from the high voltage winding of the power transformer to unidirectional current. Practically all modern receivers utilize the unilateral conductivity characteristic of the vacuum tube for this purpose

The theory of operation of a vacuum tube as a rectifier has been described in Chap X The fundamental principle is the same as that employed by Fleming in his early experiments. A circuit of a typical half-wave rectifier is shown in Fig. 141(a). A multiple secondary transformer is utilized in this circuit simultaneously to provide filament current for the heater and to supply the high voltage to be rectified to the plate. When a sine wave of alternating voltage is applied to the plate filament circuit of this system (through the load $R_L$), the plate is alternately positive and negative with respect to the filament. On the positive alternations, the electrons emitted from the filament are attracted by the positive charge of the plate, and current flows through the tube and, consequently, through the load. On the intervening alternations, the tube plate is negative with respect to the filament, and electrons from the filament are repulsed by the plate. No current flows through the tube or through the load during these periods. Since the plate is positive only on every other alternation of the input cycle, the output is in the form of pulses, as shown in Fig. 141(b). Inasmuch as the voltage on the plate

is not constant during the alternation but varies in the form of the original sine wave, the current flowing through the tube also varies in amplitude in accordance with the plate voltage. Accordingly, the output pulses are in the form of an alternation of a sine wave. Nevertheless, the output current is *unidirectional* and is therefore a direct current.

The output of a half-wave rectifier system, as shown in Fig. 141(b), can be filtered by means of an inductive-capacitive network, which will be described in the following section. The variations in amplitude of this direct current are removed by such a filter system, resulting in an output that is essentially uniform. This system is graphically depicted

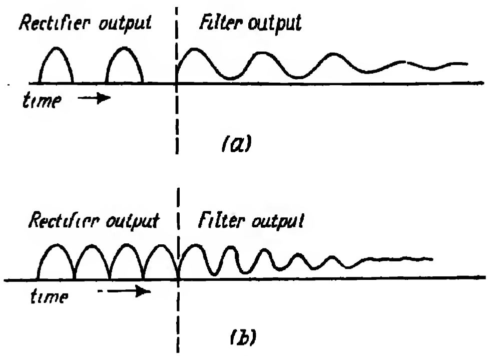



FIG 142. Graphs showing filtering action. (a) Half wave. (b) Full wave

in Fig 142(a). The d-c voltage value obtained from such a system is appreciably less than the effective value of a-c voltage originally applied to the rectifier that is, the voltage of the high voltage secondary winding. For this reason half-wave rectifiers are not widely used, and, in addition, such a system is difficult to filter properly owing to the periods of zero voltage from the rectifier.

A more efficient system in almost universal use is the full-wave rectifier system. This circuit utilizes *two* vacuum tubes so connected that each tube passes current during alternate alternations of the input cycle. Thus, if the plate of tube 1 is positive during the first alternation of input voltage, the plate of tube 2 is negative. Hence, current flows through tube 1 and not through tube 2. On the following alternation, tube 1 is negative, and tube 2 is positive. Therefore, current flows through tube 2 and not through tube 1. Current *always* flows in the output circuit throughout every part of the cycle and is always in the same direction.

Figure 143(a) is a diagram of a typical full-wave rectifier, and Fig. 143(b) shows the input and output wave forms for this circuit. The negative load connection is made to the center tap of the high-voltage secondary winding. At any instant, one end of this winding is positive

with respect to the center tap, and the other end is negative with respect to the center tap. Thus, if the end of the winding connected to the plate of tube 1 is positive during one alternation with respect to the center tap, then the end connected to the plate of tube 2 is negative with respect to the center tap. Since the center tap is connected to the filaments of both tubes through the load resistance, this makes the plate of tube 1 positive with respect to its filament and the plate of tube 2 negative with respect to its filament. Current therefore flows through tube 1. No current flows through tube 2, because of the negative charge of its plate.

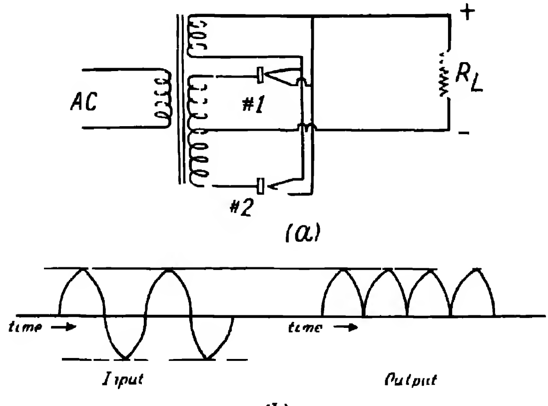



FIG. 143 (a) Full-wave rectifier circuit (b) Wave forms for circuit of (a)

The path of the electron flow is from filament to plate of tube 1, through the *upper half* of the high-voltage transformer winding to the center tap, from center tap to the negative load resistance connection, *through* the load resistance to the positive terminal, and thence back to the filament of tube 1.

On the succeeding alternation, the plate of tube 1 is negative with respect to its filament, and the plate of tube 2 is now positive with respect to its filament. Current therefore flows through tube 2. The path of this electron flow is from filament to plate of tube 2, through the *lower half* of the high voltage transformer winding to the center tap, from center tap to the negative load resistance terminal, *through* the load resistance to the positive terminal, and thence back to the filament of tube 2. In both cases, it will be noted, the current flow through the load resistance is *in the same direction*. Figure 143(b) shows the input and output wave forms for a full-wave rectifier system.

Tube manufacturers soon recognized the need for simplifying rectifier systems. The so called full-wave rectifier tube was developed especially for application in full-wave rectifier systems, such as the one described above. This tube performs the functions of two half-wave rectifier tubes

in one envelope. Two separate plates are utilized in the tube operating from a common filament or emitter. Figure 144(a) illustrates the construction of a typical full-wave rectifier tube. Typical modern tubes of this type are represented by the type 80, 5U4G, 5Y3G, 5Z3, 6X5GT, 25Z5, 35Z5-GT, and many others. Figure 144(b) illustrates a full-wave rectifier circuit utilizing a type 80 rectifier tube.

The development of the full-wave rectifier tube eliminated one of the advantages of the half-wave rectifier system over the full-wave rectifier system, namely, the need for one tube instead of two. The output of a

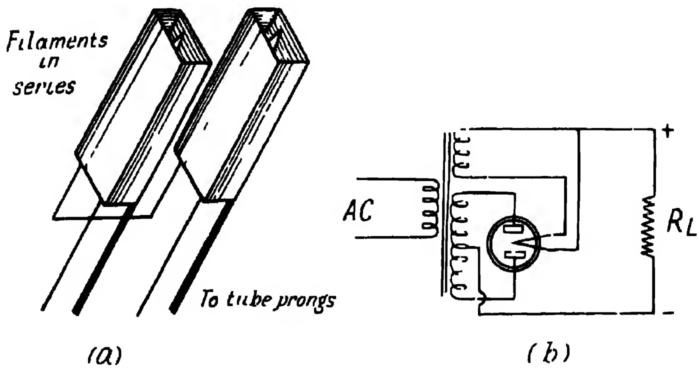



FIG. 144. (a) Internal construction of full-wave rectifier tube. (b) Rectifier circuit utilizing single full-wave tube.

full-wave rectifier is more easily filtered because of the absence of prolonged periods of zero voltage. In addition, since there are more voltage peaks per unit time than in the half-wave system, the full-wave rectifier is capable of delivering greater output for a given input. Figure 142(b) graphically depicts the wave forms of a full-wave rectifier output as the current passes through a filter network.

A unique scheme that was developed to increase the voltage output of a full-wave rectifier is shown in Fig. 145(a). The d-c voltage output of this circuit is approximately twice that obtainable from a half-wave rectifier operated on the same a-c voltage supply. In Fig. 145(a) two diodes (half-wave rectifiers) are so connected to two condensers that one diode is reversed electrically with respect to the other. Each alternation of the a-c supply cycle is therefore rectified. During the period that one tube is rectifying, the condenser across the other diode is discharging through the load and the conducting diode. Consequently, the voltage across the load is the sum of the d-c output voltage of the conducting tube and the discharge voltage of the condenser, a total voltage approximately twice the d-c voltage obtainable from a half-wave rectifier. This circuit is therefore called a **voltage doubler**. This has all the advantages of any full-wave rectifier system: easier filtering, increased output, and

so on. The type 25Z5 tube was especially designed for voltage-doubler requirements. It incorporates two separate diodes of the heater-cathode type in a single envelope. Figure 145(b) illustrates a voltage-doubler circuit utilizing a single type 25Z5 tube.

**The Filter.** The pulsating voltage delivered by the rectifier output can be smoothed into a steady d-c voltage suitable for applying to the plate of a vacuum tube by being passed through an electrical network

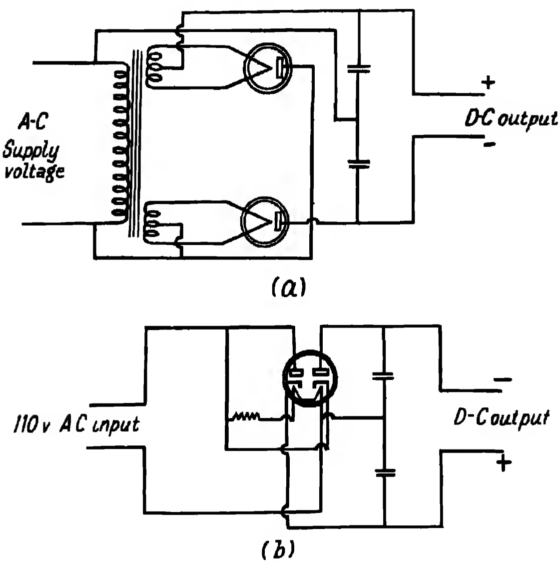



FIG. 145. Typical voltage-doubler rectifier circuits. (a) Voltage-doubler circuit utilizing two diode rectifier tubes. (b) Voltage doubler circuit utilizing a type 25Z5 full wave rectifier tube. This tube was especially developed for service in transformerless receivers.

called a **filter.** A filter is ordinarily composed of a combination of series inductances and shunt capacitances. Filters may be divided into two general classes—inductance input filters and condenser input filters—depending upon whether a series inductance or a shunt capacitor is the first unit on the input side of the filter network. The above two classifications may be further subdivided according to the number of sections or filter elements involved. A single filter section is customarily regarded as being composed of one series inductance in combination with one shunt capacitor. The most commonly used filter circuits are those shown in Fig. 146.

The fluctuation in the unfiltered output current of a rectifier is customarily called a **ripple,** since it is regular and periodic. The *frequency* of this ripple depends upon the nature of the rectifier. Thus, the output of a half-wave rectifier contains 60-c ripple (see Fig. 141(b)). The output of

a full-wave rectifier contains 120-c ripple (see Fig. 143(b)). Harmonics of the ripple frequency are always present in the rectifier output, but they are filtered more than the lower ripple frequency.

The permissible amplitude of ripple in the d-c supply of a receiver is very small and varies greatly with the type of circuit. Ripple in the power supply of a receiver is evidenced in the output in the form of a 60- or 120-c hum. If the receiver has many stages of amplification, the hum present in the first stage is amplified many hundreds of times and may assume serious proportions in the output. For this reason, the percentage of ripple must be kept very low in such receivers. On the other hand, in small midget receivers with low current drain and imperfect

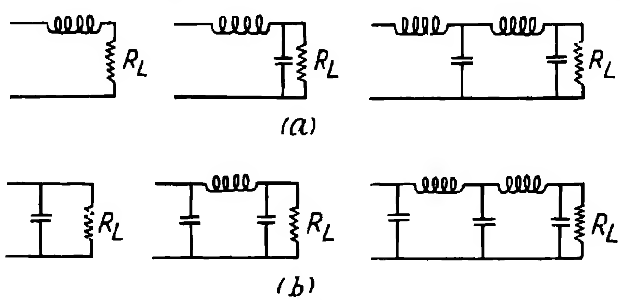



FIG. 146 A number of commonly used filter circuits. (a) Choke input. (b) Condenser input.

speaker baffling, a surprisingly large amount of ripple may be tolerated. In general, however, the percentage of ripple present in well-designed power supplies is kept to less than $\frac{1}{2}$ per cent, and values of less than $\frac{1}{10}$ per cent are quite common.

The filtering efficiency of a capacitor or a choke (inductance) is a function of the capacitive reactance and inductive reactance of these units. Generally speaking, the higher the inductive reactance of a choke, the more efficient its filtering action. Similarly, the lower the capacitive reactance of a capacitor, the more efficient its filtering action. Since inductive reactance increases with the frequency and capacitive reactance decreases with the frequency, it follows that a given choke and capacitor combination will more efficiently filter a 120-c ripple than a 60-c ripple. Consequently, a full-wave rectifier system requires less inductance and capacitance to filter it efficiently than does a half-wave rectifier system.

The filter action of an inductance is due to the counter emf of self-induction previously discussed in Chap. VII. Lenz's law states that the back emf of self-induction is always in such a direction as to oppose *any change* in the current that produces it. Accordingly, when a pulsating direct current such as the output of a rectifier is passed through an inductance, the effect of the back emf is to *decrease* the amount of

variation in the current. If the inductance is made large enough, it would be theoretically possible to eliminate entirely *all* variations in current amplitude, thus producing a uniform d-c output. Practically, however, a single inductance large enough to accomplish good filtering would be unfeasible. Therefore, the desired result is accomplished by a combination of inductance and capacitance or several such combinations.

The action of capacitor charge and discharge has been previously discussed in Chap. VIII. When the fluctuating output voltage of a rectifier is applied to a capacitor, the capacitor charges up to the peak value of the fluctuations. When the voltage input (rectifier output) to this charged capacitor tends to *decrease*, the capacitor *discharges* into the circuit. When the voltage input tends to *increase*, the capacitor charges up again. The energy consumed in charging the capacitor on the voltage increases is taken from the rectifier and causes the peaks of the pulsating rectifier output to become flattened. During voltage decreases, the capacitor discharges and this energy is returned to the circuit, thus filling in spaces between peaks in the rectifier output wave form The circuit voltage, therefore, after having been applied to a capacitor, no longer rises to sharp peaks at the crest of each pulse, nor does it fall to zero between pulses.

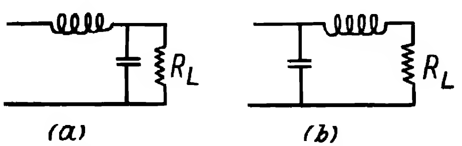



FIG. 147 (a) Elementary choke-input filter circuit. (b) Elementary capacitor-input filter circuit.

In an inductance input filter system, such as that shown in Fig. 147(a), the input inductance prevents any rapid changes (either increases or decreases) in current through it. Therefore, the variations of the voltage impressed across the capacitor are not so great as were those of the rectifier output, and the capacitor charges up to the *average* value of the original input fluctuations. In the capacitor-input filter system of Fig. 147(b) the capacitor is connected directly across the output of the rectifier, and the capacitor charges up to the *peak* value of the voltage fluctuations. The larger the capacitance of this capacitor, the more charge it can take the more it must discharge to have its voltage drop to a certain point, and the longer it takes to discharge to this point. Accordingly, in a capacitor input filter, the voltage output increases as the capacitance of the input capacitor is increased.

In the choke input filter, the size of the capacitors used does not appreciably affect the magnitude of the output voltage. The capacitance values do, of course, affect the filter action. The voltage across any capacitor of a choke input filter is *always* lower than the voltage applied across the input capacitor in a capacitor input filter operating from the same rectifier source. In the latter case, the voltage applied to the capacitor is very nearly the peak voltage of the rectifier output. For this

reason, the voltage output of a capacitor-input type of filter is always higher than that of a choke-input type of filter. In the choke-input circuit, the voltage applied to the first capacitor is equal to the *average* value of the input voltage less the *IR* drop in the first inductance, which is due to the resistance in the first inductance.

The number of sections utilized in either type of filter is dictated by the ripple voltage permissible and by economic factors. Thus, a filter system could be designed for a receiver for any given percentage of ripple voltage utilizing only one section. If the permissible minimum ripple voltage is extremely small, however, the physical size of inductance and capacitance necessary to achieve the desired result would be so great as to become impractical. Thus, if a large inductance is necessary, that is, one having many turns of wire, then wire of fairly heavy gauge must be used to prevent undue voltage drop in the filter due to the resistance of the wire. The desired results are obtained by utilizing a filter having two, or even more, sections, and in such a filter, much smaller components may be used with the same result. Filters having more than two sections are very rarely necessary in receiver circuits.

It is possible to design filters mathematically with a fair degree of accuracy to give any desired attenuation of ripple. It is customary in such mathematical treatment to consider each section of a filter separately. Accordingly, the circuit constants for the input section are treated first; then, the circuit constants of the subsequent sections in their order are treated.

In a capacitor-input filter, the input capacitor is considered as the first section of the filter. The first series inductance and second capacitor are considered jointly as the second section. All the subsequent sections are then treated in the same manner as the second section.

In a choke input filter, the first section is taken as the input choke and first capacitor. All subsequent sections are treated in like manner.

The correct value of capacitance for the input section of a capacitor input filter is computed from the formula

$$C = \frac{1.41}{2\pi fRk}, \qquad (15)$$

where $C$ = capacitance of input capacitor in farads,
$1.41$ = constant for all cases,
$\pi$ = the constant 3.14,
$R$ = resistance of the load to which the filter is to be coupled, plus d-c resistance of any series inductances;
$f$ = the cutoff frequency (minimum frequency to be attenuated);
$k$ = ripple-frequency percentage output of this section in decimal form (standard practice is to bring the output-ripple percentage down to less than 10 per cent in the first section).

The cutoff frequency $f$ is the minimum frequency to be attenuated. In filtering the output of a full-wave rectifier, this frequency (in the case of a 60-c a-c supply) will be 120 c; in other words, 120 c will be the fundamental ripple frequency. There will be no lower ripple frequencies present in the rectifier output, since all harmonics of this fundamental frequency will be multiples of 120. If the output of a half-wave rectifier is being filtered, the fundamental ripple frequency will be 60 c, and the cutoff frequency to be substituted in Eq. (15) will, accordingly, be 60 c.

The calculation of the circuit constants for the second section of a capacitor-input filter is based on the formula

$$LC = \frac{a - b}{b4\pi^2 f^2}, \tag{16}$$

where $LC$ = product of series inductance in henrys and shunt capacitance in farads;
$a$ = percentage input ripple voltage;
$b$ = percentage output ripple voltage;
$\pi$ = constant 3.14;
$f$ = cutoff frequency.

The percentage values of ripple voltage should be substituted directly in Eq. (16). They should *not* be first reduced to decimal values. Thus, if there is 10 per cent input ripple voltage and it is desired to reduce the ripple to 0.4 per cent (four tenths of 1 per cent), the value of $a - b$ in Eq. (16) becomes 9.6. The value of $b$ in the denominator will be 0.4. The percentage input ripple voltage to this section is obtained from the percentage of output ripple voltage calculated for the first section from Eq. (15).

Circuit constants for subsequent sections in this same filter—third, fourth, and so on—are obtained by substitution in the same Eq. (16). The percentage input ripple voltage in each case will equal the percentage output ripple voltage of the preceding section.

The $LC$ product obtained from the solution of Eq. (16) can be derived from an infinite number of combinations of $L$ and $C$. Considerable leeway is available in the choice of these individual circuit elements, the proper selection being restricted only by practical circuit considerations and design economy.

A very popular form of the capacitor-input type of filter is the *brute-force* filter, so named by Ballantine. This filter consists of two-sections and is shown in Fig. 148(a). The design procedure for this type of filter is illustrated in the following problem.

**Problem.** It is desired to design a brute-force type of capacitor-input filter to supply 260 v at 65 ma direct current from a full-wave rectifier. The frequency of the single-phase supply source is 60 c. The output ripple voltage must not exceed 0.1 per cent (one tenth of 1 per cent).

**Solution.** The effective resistance of the load ($R_L$, Fig. 148) will be

$$R_L = \frac{260}{0.065} = 4{,}000 \text{ ohms.} \tag{17}$$

Standard practice is to keep the percentage output ripple from the first section down to less than 10 per cent. The percentage input ripple to the second section is therefore chosen as 10 per cent, and the percentage output ripple from the section to satisfy the problem requirements must be 0.1 per cent. For a 60-c supply, the fundamental ripple frequency will be 120 c.

Substituting in Formula (16),

$$LC = \frac{10 - 0.1}{0.1(4)(\pi)^2(120)^2} = 0.000174, \tag{18}$$

where $L$ is in henrys and $C$ is in farads. For convenience in handling, this can be converted to 174 by multiplying by 1,000,000. $L$ can then be taken in

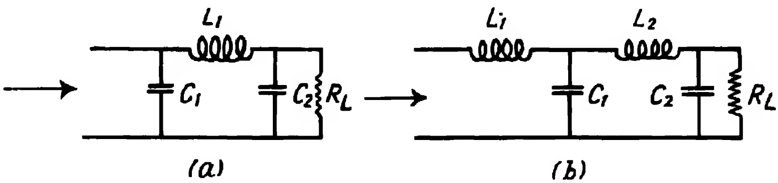



FIG. 148.

henrys and $C$ in microfarads. An 8-$\mu$f capacitor and a 22-henry choke will fulfill the requirements nicely. The values of the actual elements chosen for this section need not be exact, so long as they are not *smaller* than the calculated requirements. Thus, a 23- or 25-henry choke could be used without detrimentally affecting the output ripple voltage. Any discrepancy in calculation resulting from such a choice will tend to *decrease* the output ripple voltage, so that in any case, if the calculated minimum requirements are met, the net output ripple voltage will be 0.1 per cent or less.

It is important to choose a choke that has a minimum of d-c resistance so that excessive voltage drop will not be experienced in the filter at the required current flow. If the choke having a resistance of 300 ohms is chosen for the filter in question, this value must be added to the load resistance computed in Eq. (17) to find the effective load resistance across the first section of the filter. By substituting in Formula (15),

$$C = \frac{1.41}{6.28(120)(4{,}300)(0.1)}, \tag{19}$$

where 120 = fundamental ripple frequency;
4,300 = effective load resistance (4,000 + 300);
0.1 = ripple output from this section in decimal form (10 per cent).

Solving,

$$C = 0.00000445 \text{ f}, \tag{20}$$

or

$$C = 5\ \mu\text{f}. \tag{21}$$

A brute-force filter to meet the given requirements would therefore have circuit elements as follows:

$$L_1 = 22 \text{ h}, \tag{22}$$
$$C_1 = 8\ \mu\text{f}, \tag{23}$$
$$C_2 = 5\ \mu\text{f} \tag{24}$$

The calculation of circuit elements for a choke-input filter requires slightly different treatment. The input choke requires special consideration, since satisfactory peak current and the degree of voltage regulation depend upon its inductance value. The circuit values for the first section of a choke-input filter are calculated from the formula

$$LC = \frac{0.667}{k4\pi^2 f^2}, \tag{25}$$

where $k$ = ripple output percentage from this section in decimal form;
$f$ = cutoff ripple frequency (fundamental ripple frequency);
0.667 = a constant for full wave 60 c power supply.

The value obtained for $LC$ in the above formula is the product of $L$ in henrys and $C$ in farads. The exact value of $L$ is derived from the following formula:

$$L = \frac{0.667R}{2\pi f}, \tag{26}$$

where 0.667 = a constant for full-wave 60-c supply;
$R$ = effective load resistance for the first section, including the filter output load plus the series resistances of the input inductance and any subsequent inductances in the filter;
$f$ = cutoff ripple frequency [for the full-wave 60-c source in question, $f$ will be 120 c in both Formulas (25) and (26)].

The calculations for any subsequent sections of a choke-input type of filter are computed in the same manner as the subsequent sections in a capacitor-input type of filter. They are based on Formula (16).

**Problem.** It is desired to design a two-section choke-input type of filter, as shown in Fig. 148(b). The filter is to operate from a full-wave rectifier on a 60-c single-phase source. The power to be supplied is 200 v at 50 ma. The permissible output ripple voltage is 0.4 per cent (four tenths of 1 per cent).

**Solution.** For the full-wave rectifier under consideration, the fundamental ripple frequency will be 120 c. According to standard practice the ripple output from the first section will keep down to 10 per cent. The input ripple voltage to the second section will, therefore, be 10 per cent. Substituting in Formula (16),

$$LC = \frac{10 - 0.4}{0.4(39.44)(120)^2} = 0.0000422, \tag{27}$$

where $L$ is in henrys and $C$ is in farads. To facilitate handling, this figure can be converted to 42.2 by multiplying by 1,000,000. $L$ is then taken in henrys and $C$ in microfarads. A 10-h choke and a 4.22-$\mu$f capacitor would meet the exact computed requirements. But since 4.22 $\mu$f is an odd capacitance value, 5 $\mu$f may be used, and the minimum requirements will be satisfied.

The circuit-element values for the first section are found by substituting in Formula (25). Thus,

$$LC = \frac{0.667}{0.1(39.44)(120)^2} = 0.0000117, \tag{28}$$

where $L$ is in henrys and $C$ is in farads. The exact values of $L$ and $C$ are found by substituting in Formula (26). The output load resistance is found by substituting the given values of voltage and current in Ohm's law. Thus,

$$R = \frac{E}{I} = \frac{200}{0.05} = 4{,}000 \text{ ohms} \tag{29}$$

If the chokes chosen for service in both sections are assumed to have a total resistance value of 200 ohms, the net, or effective, load resistance into which the first section is working will equal

$$4{,}000 + 200 = 4{,}200 \text{ ohms} \tag{30}$$

Substitution in Formula (26) gives

$$L = \frac{0.667(4{,}200)}{6.28(120)} = 3.7 \text{ h}. \tag{31}$$

Since the value obtained in Eq. (28) is the product of $L$ and $C$ for this section, the value of $C$ is found by dividing this product by $L$.

$$C = \frac{0.0000117}{3.7} = 0.0000032 \text{ f} = 3.2 \ \mu\text{f} \tag{32}$$

An inductance value of 4 h and a capacitance of 4 $\mu$f would, therefore, ensure an output ripple voltage not in excess of 0.4 per cent for the entire filter.

The capacitors used in filter circuits must be capable of continuously withstanding a d-c voltage equal to the peak a-c voltage applied to the rectifier. Often it is found necessary to use two capacitors in series to take the place of a single high voltage rating capacitor. This practice is not recommended for any power pack subject to continuous or semicontinuous service. When capacitors are connected in series, the d-c voltage is divided between them in proportion to their leakage resistances rather than to their dielectric strength. Hence, the voltage distribution is variable and uncertain. When capacitors are connected in this way, it is desirable to shunt each of them with an appropriate resistor to help stabilize the voltage distribution across them.

The inductances used in a filter system are of the laminated iron-core type. Cores made of magnetic material, such as iron, greatly increase the flux produced by a given magnetomotive force. Consequently, by the use of magnetic core material it is possible to obtain a large inductance

in a physically small unit, which utilizes relatively few turns of wire, with all the attendant advantages. Iron cores, however, have a serious disadvantage particularly important in filter circuits, in that they are subject to the phenomenon of *magnetic saturation.*

The magnetic permeability of iron may be compared to the ability of a sponge to absorb water. When there is very little water in the sponge, it will very readily soak up more, when the sponge is almost saturated, however, it soaks up additional water only with difficulty. Similarly, when the current flowing through an iron-core inductance is very low, a comparatively small increase in current will appreciably increase the flux density. When the current flowing through the iron-core inductance is high, the flux density approaches an upper limit, and a similar small increase in current will increase the flux density very little.

When the current is so high that the flux density is very near its upper limit, the core is said to be **magnetically saturated.** The incremental permeability of the iron, which is the ratio of the change in flux density to a given small change in magnetomotive force, decreases as the current in the windings increases. Since the apparent inductance for small changes in current, or ripple, depends on the incremental permeability, the effectiveness of an inductance in a filter will decrease as the load current from the power supply increases.

In order to obtain a larger inductance at large current values, common practice is to provide an air gap in the core of the choke. Such a choke has a lower inductance at low current values than has a similar choke with no air gap, and it also has a higher inductance at large current values.

This change in inductance with magnetomotive force is sometimes put to good use, as in the *swinging choke*, which utilizes a small air gap in its core. The swinging choke is used only as the first inductance in a choke input filter. At lower current values the percentage of ripple is somewhat higher, but the lower inductance of the input choke allows the filter to function more like a capacitor input filter. The result is a power supply with better regulation, that is, more-constant voltage under varying load.

The application of voltage divider and bleeder circuits for power packs has been discussed in Chap. IV and need not be repeated here. A complete power-supply unit, or power pack, includes power transformers, rectifier, filter system, and voltage divider.

Rectifier tubes for use in power packs are rated according to maximum permissible effective a-c voltage per plate, maximum current-carrying capacity, and *maximum peak inverse voltage*. The last expression refers to the maximum negative voltage that the tube can withstand without breakdown (arcing between electrodes). The inverse voltage present at voltage peaks across a rectifier tube is usually well in excess of the normal

circuit voltages. In general, for the type of rectifier circuits previously discussed, the maximum peak inverse voltage is 3.14 times as great as the d-c output voltage from the rectifier.

## *THE DETECTOR*

The heart of any receiving system is the detector circuit where the received r-f signals are converted to a-f currents. Before proceeding with the theory of detection, it is necessary that the reader be familiar with the nature of the complex wave form which constitutes the received signal.

The theory of modulation is discussed in detail in Chap. XIII. The modulated wave radiated by a radiotelephone transmitter will not,

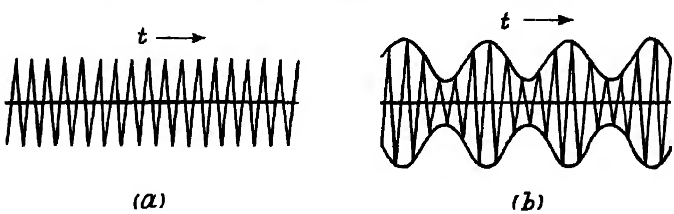



FIG. 149

therefore, be analyzed objectively here. It will suffice for the purpose of this section to consider the nature of the modulated wave without regarding the manner of its creation.

The signal voltage present in a receiving antenna system in the case of a pure, unmodulated signal is in the form of a symmetrical h-f (r-f) alternating current as shown in Fig. 149(a). It will be observed that the amplitudes of successive peak values of this alternating current are the same. Such a signal voltage is therefore said to be **unmodulated.***

The process of modulation consists basically of superimposing an a-f current upon such an r-f current. An r-f current modulated by a single a-f current having a sine-wave characteristic is shown in Fig. 149(b). Such a signal is said to be **amplitude-modulated** because the process of modulation has changed the amplitudes of successive peak values of the original r-f signal. The peak values of r-f voltage vary in exact accordance with the amplitude of the impressed a-f voltage. The r-f signal voltage is called **carrier** because it fulfills the function of "carrying" the audio voltage. When a number of different audio frequencies are superimposed upon the carrier simultaneously, as occurs with voice modulation, it is readily apparent that the resulting modulated wave form is quite complex.

The primary function of a receiver is to separate the audio component, or *envelope*, as it is called, from the carrier, or r-f component. Offhand,

* Frequency modulation is discussed separately in a later section of this chapter.

it would appear that this could be quite simply accomplished by impressing the modulated signal across a circuit that is resonant only to the audio frequency. If this were done, however, the resulting a-f voltage would have the form shown in Fig. 150(a). *Two* audio frequencies would be present in the circuit, one above the zero line and one below the zero line. An inspection of Fig. 150(a) shows that at any instant the audio voltage above the zero is equal in amplitude and opposite in polarity to the audio voltage below the zero line. The two voltages would, consequently, balance out, resulting in zero audio voltage. Obviously, it is necessary to *rectify* the modulated carrier before impressing it across an audio circuit. This process of rectifying a signal voltage is called

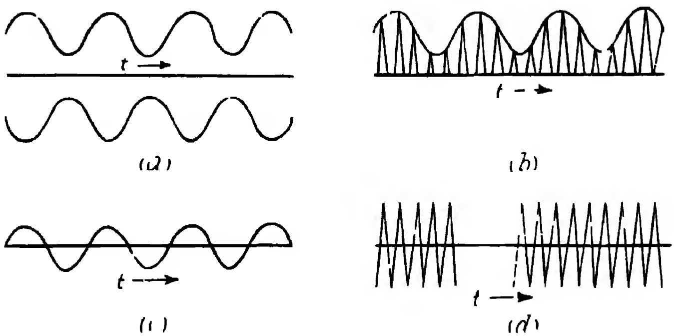



Fig. 150

"detection" in receivers. The detected or rectified, output of a detector circuit having an input in the form shown in Fig. 149(b) would appear in the form shown in Fig. 150(b). If this voltage output is now impressed across an amplifier that responds only to audio frequencies, the resultant output would appear in the form shown in Fig. 150(c). If this pure a-f signal shown in Fig. 150(c) is then applied to a sound-converting unit, such as an earphone, as described in Chap. XIV, it will set up sound waves of the same frequency as the audio voltage.

Although numerous systems of detection have been devised since the inception of radio, the most important are those utilizing the vacuum tube. Modern receivers employ three major systems of vacuum-tube detection: *plate detection*, *grid-leak detection*, and *diode detection*.

**Plate Detection.** A plate detector is essentially a system in which the grid bias is kept at a negative value high enough to operate the tube at or near the cutoff point. Since the tube depends for its detector action upon the critical value of grid bias, this system is often called **grid-bias detection.** Figure 151 is a graph of the input and output wave forms for a plate-detector tube.

One of the principal advantages of plate detection is that relatively large a-f power outputs may be obtained with a given input signal

voltage because plate detectors amplify as well as detect the signal voltage. This system of detection is used where relatively large values of input signal voltages are handled, as in receivers where considerable amplification of the signal is accomplished prior to detection. The plate detector is capable of handling large inputs because of the high value of bias that is used. A considerable value of signal voltage would have to be applied before the grid would be driven positive.

Plate detection is not suitable for low detector input applications. At low values of input signal voltage, the sensitivity of a plate detector is

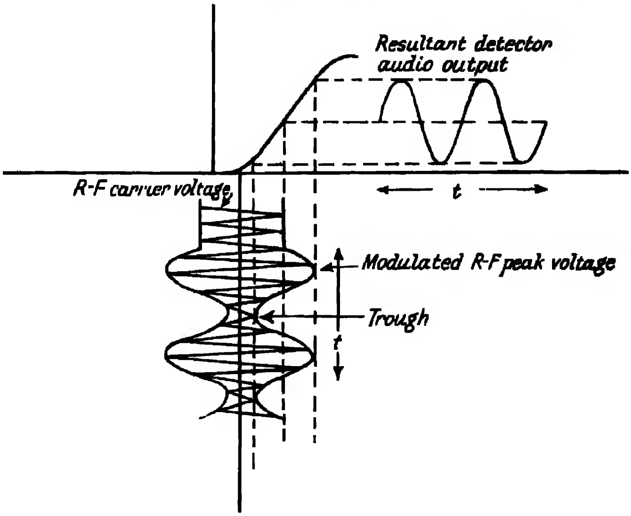



FIG. 151

lower than in other forms of detector. In addition, because of the curvature of the characteristic at low input values, considerable distortion of the signal occurs. Nevertheless, the major portion of a plate detector's characteristic is linear, and good fidelity and output may be obtained by its use.

**Grid-leak Detection.** In the grid-leak system of detection, the entire process of detection takes place in the grid circuit, and the plate circuit functions merely as an amplifier. In this system of detection, no external bias is applied to the tube (see Fig. 152). The modulated carrier is applied to the grid-cathode circuit through capacitor $C_1$. Since the a-f output in a grid detector depends entirely upon the amplitude of the a-f component across the grid-cathode input, it is necessary to build up maximum a-f voltage across this circuit. Therefore, the value of $C_1$ is chosen to give a very high value of reactance at audio frequencies while offering a minimum of reactance to the r-f carrier. Values of 0.0001 and 0.00025 $\mu$f are common.

In the absence of bias, the grid becomes positive on positive alternations of signal voltage, and electrons from the cathode are attracted to the grid. Unless provision were made to allow a direct current to flow from grid to cathode to restore these electrons to the cathode, the electrons would accumulate on the grid. In a short time the negative charge on the grid would become so great that the tube would be "blocked," that is, no plate current would flow. This electron accumulation is prevented by introducing a resistor across $C_1$ ($R_g$, Fig. 152), thus permitting a direct current to flow around the grid-cathode circuit. This resistor is called a **grid leak,** since it fulfills the function of allowing the grid electrons to leak back to the cathode. The value of grid leak is kept fairly high (usually from 1 to 3 megohms) in order not to short-circuit the high reactance of condenser $C_1$ to audio frequencies. Nevertheless, a compromise value must be attained that will not cause distortion at higher amplitudes of input voltage. The greater the input voltage, the greater the current flow through the grid leak and, hence, the greater the average negative bias on the grid. When the grid bias increases, the effective resistance of the grid circuit increases, causing possible distortion at the higher audio frequencies. The grid leak should not be permitted to exceed a value that will introduce a-f distortion.

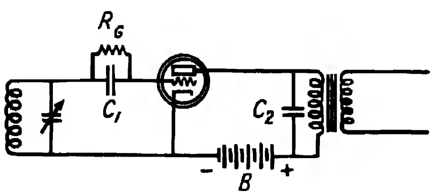


FIG. 152 Grid-leak detector circuit.

In the plate circuit of the grid leak detector, the r-f component of the rectified plate current is by passed across the output by a capacitor $C_2$. $C_2$ is given a value that will present minimum reactance to radio frequency and yet offer high opposition to audio frequencies. Practically none of the r-f component, therefore, appears across the input to the succeeding audio amplifier, and the a f component is comparatively unhindered.

Grid-leak detectors find their greatest application in circuits where extreme sensitivity is of paramount importance. They are easily overloaded with resultant distortion and, in general, should not be used in circuits where amplification is available ahead of the detector.

Grid-leak detectors are especially suitable for regenerative-detector circuits where it is desired to take advantage of the extra amplification and sensitivity afforded by feedback. Such circuits are also used when it is desired to make the detector oscillate to facilitate the reception of c-w (telegraph) signals, which are unmodulated. Telegraphy is accomplished by interrupting the r-f carrier so that the emitted r-f pulses form the dots and dashes of the Morse code. Figure 150(d) illustrates the signal-voltage transfiguration for the letter $A$ in Morse code. Detecting

such a signal in itself accomplishes no useful purpose, since there is no a-f component present to appear in the output. The principle of heterodynes discussed in Chap. XI is applied to generate the necessary a-f component. The feedback in a regenerative detector is increased sufficiently to make the detector oscillate. By slightly detuning the input circuit, the frequency of oscillations is made to differ from the frequency

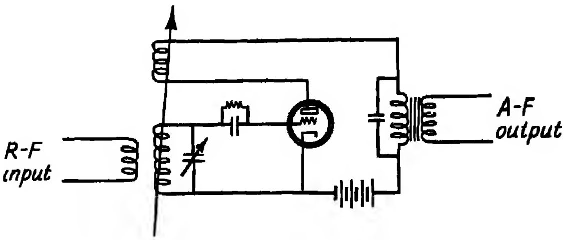



FIG. 153 Regenerative detector circuit.

of the signal voltage by an amount that is an audio frequency. The resulting difference beat frequency, or heterodyne, is then present in the complex signal impressed upon the detector input circuit and appears as an a-f component in the plate circuit. This function is also performed by a separate oscillator tube coupled to a nonregenerative detector input circuit. A typical regenerative detector circuit is illustrated in Fig. 153.

**Diode Detection.** Because of distortion and hum arising in the audio-amplifier stages following the detector, it has become standard practice to use considerable amplification preceding the detector and comparatively little in the audio amplifier. Such procedure necessitates a plate detector capable of handling high input voltages and delivering high audio power output. The distortion that often occurs in such detectors because of their nonlinearity led to the development of the diode detector. The diode detector operates on the same principle as the half-wave rectifier utilized in power packs and has the advantage of true linearity. An ordinary triode tube, such as the type 56, is often used as a diode detector, as shown in Fig. 154. When the positive half cycles of input signal voltage are applied to the grid, current flows in the grid-cathode circuit through resistor $R_1$. The $IR$ drop across $R_1$ supplies the audio voltage to be impressed across the following audio stage. During negative half cycles, no current flows through the grid circuit, and true rectification is accomplished.

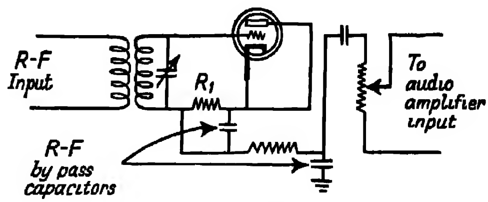



FIG. 154. A diode detector circuit employing a triode tube

Modern design practice led to the development of special tubes for diode detector circuits, for example, the type 55, where a diode detector and triode audio amplifier are combined in the same envelope.

Although diode detection is not inherently a sensitive system, it possesses the tremendous advantages of minimum distortion and the ability to handle either high or low input voltages with almost the same degree of linearity.

## THE R-F AMPLIFIER

Radio-frequency amplifiers may be divided into two broad classifications—*untuned* and *tuned*. Untuned r-f amplifiers are those in which any or all the r-f signal voltages existing in the antenna system are

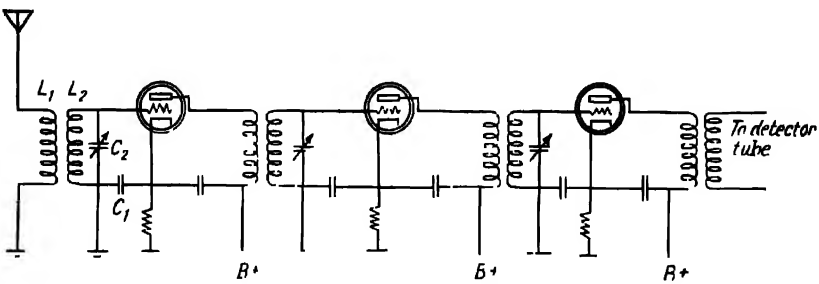



FIG. 155 A typical triode tuned r-f amplifier

amplified (theoretically equally) before being impressed across the detector input circuit. In receivers using such a system discrimination against undesired signals is accomplished solely by means of the selectivity of the detector input resonant circuit. Because of the relatively close spacing of radio stations in the frequency spectrum under present conditions, the selectivity afforded by a single tuned circuit is not sufficient for proper discrimination against unwanted signals. This is especially true when it is necessary to tune in weak distant stations through powerful local station interference. For that reason untuned r-f amplifiers are almost never used in modern receivers.

**The Tuned R-F Amplifier.** Figure 155 illustrates a typical three-stage tuned r-f amplifier utilizing triode tubes. Any signal voltages present in the antenna cause a difference of potential between the antenna itself and ground, thus causing a flow of current through coil $L_1$ to ground. Normally, therefore, a great many currents of various frequencies will flow through coil $L_1$, each representing a different station picked up by the antenna. All these currents are induced by transformer action to coil $L_2$, which is customarily wound on the same form. It is also common practice to provide a small amount of voltage step-up in the transformer $L_1L_2$. The coupling between these coils is usually determined empirically

and is so chosen as to provide maximum signal to unwanted noise ratio. All the above currents circulate in the tank circuit composed of the inductance $L_2$ on one side and the capacitor $C_2$ on the other. Each of the currents causes a difference of potential between the grid side of the tank circuit and the cathode circuit. The voltage across the grid-cathode, or input, circuit of the first r-f amplifier tube for each frequency is derived by Ohm's law

$$E = IZ, \tag{33}$$

where $I$ = current through $L_2C_2$ at one particular frequency;
$Z$ = impedance of this parallel circuit at that frequency;
$E$ = voltage developed across the circuit at that frequency.

It is apparent that there will be as many voltages impressed across the grid-cathode of the tube as there are currents of different frequencies. The parallel circuit $L_2C_2$, however, is resonant at only *one* of these frequencies. Operating as an antiresonant circuit, consequently, this circuit offers maximum impedance to only *one* frequency. The voltage drop across the circuit and the resultant voltage impressed across the tube will therefore be very great at this one frequency and relatively smaller at other frequencies. Since $C_2$ is a variable capacitor, this frequency can be chosen at will within the limits of the circuit constants by varying the resonant frequency of the circuit, or tuning it.

The process is repeated in the following amplifier stages. In each stage the tank circuit is resonated at the same desired frequency. As a result, a comparatively powerful signal voltage is presented to the detector at the desired frequency. The voltages at all other frequencies are so small by comparison, having obtained little or no amplification, that the detector is not sensitive enough to rectify them. Hence, they do not appear in the a-f component of the detector output.

If the parallel tank circuits of each grid of an r-f amplifier were composed of pure inductance and pure capacitance, the impedance would rise to infinity at resonance (see Chap. IX). Actually, of course, all such circuits possess resistance so that the impedance passes through a maximum value at resonance. The smaller the resistance, the higher the impedance value at resonance. If such a parallel circuit is a high-resistance circuit, the resonance impedance value is comparatively low. The ratio of impedance at resonance to impedance at frequencies other than the resonant frequency is therefore diminished. Consequently, a considerable portion of the circuit selectivity is lost.

In the usual type of resonant-frequency circuit, the r-f resistance of the circuit is almost entirely in the coil. The losses in a properly designed capacitor are negligible in comparison with those in the coil. It can be seen, therefore, that the selectivity of a parallel resonant circuit is a function primarily of the coil resistance and the coil reactance. The

efficiency of a coil in this respect is conveniently referred to in terms of the ratio of the coil reactance to the effective resistance of the coil. It is standard practice to represent this ratio by the symbol $Q$. In modern design practice, the $Q$ of a coil is considered one of the fundamental coil properties. Expressed mathematically,

$$Q = \frac{X_L}{R} = \frac{2\pi f L}{R} \quad (34$$

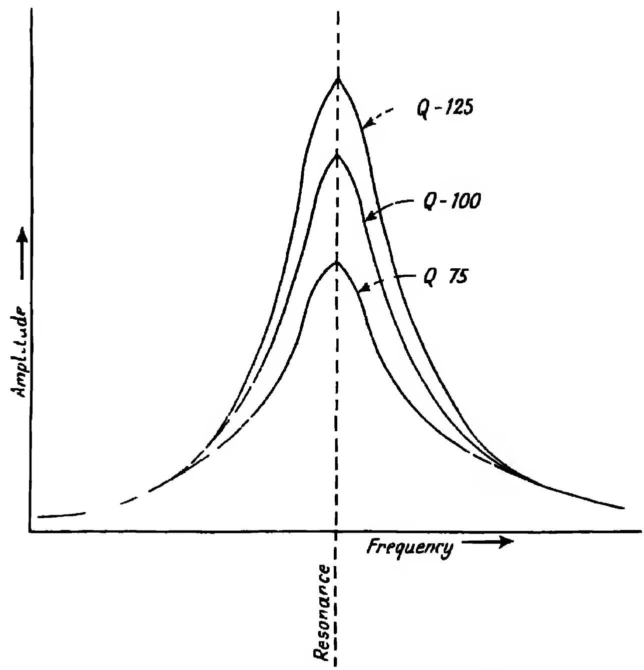



FIG. 156 Effect of circuit $Q$ on selectivity.

It is apparent that the higher the $Q$, the greater the selectivity of a parallel circuit. $Q$ values from 50 to several hundred are quite common in modern receivers. Figure 156 illustrates several frequency response curves for circuits having different values of $Q$. The amplitude of the signal voltage developed across the parallel circuits having different values of $Q$ is shown at the resonant frequency and for frequencies a considerable amount on either side of resonance.

As described in Chap. XIII, a modulated carrier signal is composed not only of the original carrier radio frequency but also of the side-band radio frequencies on either side of the carrier. These side bands are the result of the heterodyning of the carrier frequency with the various audio frequencies superimposed upon it by the process of modulation. In order

properly to transmit such a complex signal through a receiver network, the side bands must be amplified with the same fidelity as the carrier frequency. Consequently, in the design of r-f amplifiers for receivers intended for modulated carrier reception, it is necessary that the frequency response curves for the tuned circuits have flat tops over the region covering both side bands and carrier frequencies. Often considerable selectivity is sacrificed to accomplish this end. In many cases, the desired result is achieved by tuning successive circuits in an amplifier to resonance at frequencies slightly removed from and on opposite sides of the carrier frequency. This is shown in Fig. 157.

Early receivers incorporating tuned r-f amplifiers used separate variable

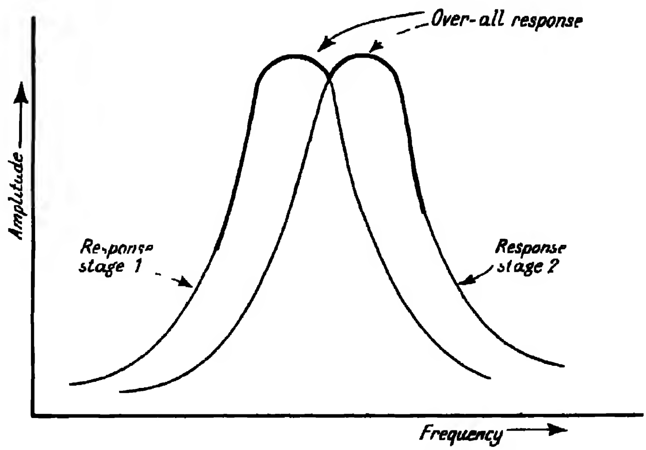



FIG. 157 Amplifier response curves

capacitors for the tuned circuit of each stage. Each stage had to be tuned separately to resonance with the desired signal, thus necessitating a number of controls for proper manipulation of the receiver. Practically all modern receivers utilize ganged capacitors. These are capacitors that are physically connected so that they may be tuned simultaneously by means of a single control. Early types of ganged capacitors utilized some form of belt or gear drive. Modern ganged capacitors are manufactured as a single unit with the rotor plates of each section mounted on a common shaft.

The principal trouble experienced with triode-type r-f amplifiers, such as that illustrated in Fig. 155, was a tendency to oscillation. As described in Chap. XI, such oscillation is likely to occur in any amplifier circuit when the plate circuit of a tube is tuned to the same resonant frequency as the grid circuit. This condition exists especially at the higher frequencies, since any capacitive coupling between input and output

circuits of an amplifier becomes lower in reactance as the frequency increases.

In a triode amplifier, practically all the feedback coupling occurs through the plate-to-grid interelectrode capacitance of the tube, but this condition may be readily circumvented by introducing an additional capacitance between output and input circuits. The additional capacitance is usually in the form of a small variable capacitor (such as a trimmer) whose capacitance is experimentally adjusted until it is exactly equal to the plate to grid interelectrode capacitance. This capacitor is so connected in the circuit that at any instant the current through it is 180° out of phase with the feedback current flowing through the plate to grid capacitance. Since the two currents are then equal and opposite in polarity the feedback current through the tube is balanced out or neutralized by the external capacitor current. This process is called **neutralization,** and tuned r-f receivers incorporating neutralized r-f amplifiers are called **neutrodynes.**

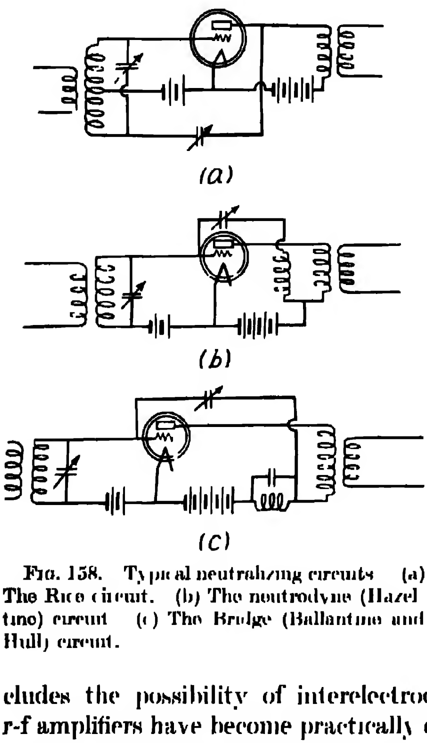


FIG. 158. Typical neutralizing circuits (a) The Rice circuit. (b) The neutrodyne (Hazeltine) circuit (c) The Bridge (Ballantine and Hull) circuit.

A number of ingenious schemes have been devised for accomplishing neutralization in r-f amplifiers and a few of the more commonly used circuits are illustrated in Fig. 158. Since the advent of the screen-grid tube, whose internal construction precludes the possibility of interelectrode feedback (see Chap. X), triode r-f amplifiers have become practically extinct in receivers Neutralization is still an important function in many transmitter circuits, however, and is discussed in detail in Chap. XIII.

**The Screen-grid R-F Amplifier.** So far as the principle of amplification is concerned, the screen grid amplifier functions in exactly the same manner as the triode amplifier. The principal advantage of screen-grid amplifiers is the lack of oscillation troubles due to interelectrode capacitance. Modern receivers utilize pentode-type tubes as r f amplifiers because of the greater permissible plate swing and resultant greater gain. Both the tetrode- and pentode-type tubes permit much greater amplification than is possible with the triode tube. In addition, greater

circuit efficiency is possible because of the better impedance match between the high plate resistance of pentode and tetrode tubes and the coupled resonant circuit plate load.

Care must be taken, because of the higher gain per stage in this type of amplifier, to reduce stray coupling to a minimum. It is usually necessary that tetrode and pentode r-f amplifiers be well shielded throughout. In addition to shielding of the tube and r-f transformers, it is sometimes necessary to shield individual leads. Feedback troubles are often experienced because of the coupling that exists by virtue of the feeding of several amplifier stages from a common power supply. In order to minimize coupling from this source, individual resistance-capacitance filters are usually inserted in the power (plate and screen grid) leads.

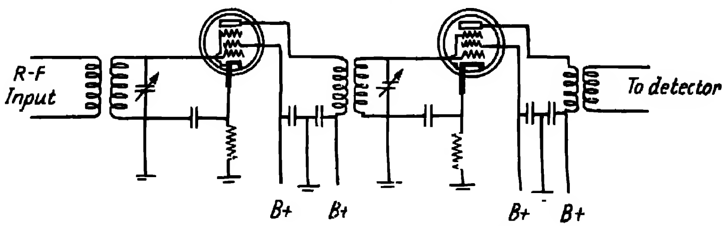


FIG. 159 A pentode r-f amplifier

Such filters operate on the same principle as filters in power packs. For reasons of space and economy, resistors are used in place of choke coils. Since the total current flowing through such a resistor in a single power lead is comparatively small, the d-c voltage drop is not out of bounds and can be tolerated. Many receivers utilize choke coils, however, if they have low d-c resistance. Since the a-c component under consideration here is an r-f component, such chokes are r-f chokes and have no core. In well designed receivers these r-f chokes are also shielded.

A number of tubes have been developed that are especially adaptable for r-f amplifier service. Such tubes as the types 58, 6D6, 6K7, 6U7G, and 6SK7 are examples. Significant among the electrical features of these tubes are the extended mutual-conductance operating range and the ability to handle signal voltages without cross modulation and modulation distortion effects. In addition, the metal types of tube, such as the 6K7, are efficiently shielded. Figure 159 illustrates a typical modern pentode-tuned r-f amplifier.

The method of obtaining the proper bias voltage to apply to the grid of an r-f amplifier tube is of interest. In general, there are several methods of obtaining this bias, but the first, and simplest, is battery bias, as shown in Fig. 160(a). In this method, bias is applied to the grid by connecting a battery of the proper voltage in the grid-cathode circuit so that the grid is always kept negative by the proper amount with respect to the

cathode. Because of limited life and other undesirable characteristics of batteries, the use of battery bias is comparatively limited. In modern practice, battery bias is confined to receivers that also require batteries for plate supply, such installations, for example, as marine (ship) receivers and many forms of portable receiving equipment.

The most popular method of obtaining bias in receiver amplifier circuits is **self-bias,** or, as it is commonly called, **cathode bias.** A resistor ($R_k$, Fig. 160[b]) is inserted in series with the cathode return circuit of

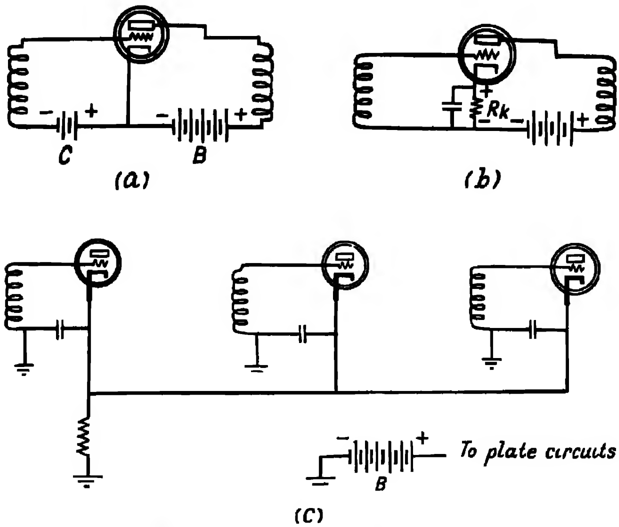



FIG. 160 Commonly used bias circuits (a) Battery bias. (b) Cathode or self bias. (c) Back bias

the amplifier tube to be biased. Since the resistor is then in series with the plate B supply circuit, it is apparent that the tube plate current flows through the resistor. A voltage drop will therefore exist across resistor $R_k$. Because of the direction of current flow, the grounded end of the resistor (Fig. 160[b]) will be negative with respect to the upper, or cathode, end of the resistor. Inasmuch as the grid is also connected to ground through the input inductance, it follows that the grid will also be negative with respect to the cathode. The exact value of cathode resistor to use for a given application is obtained from the formula

$$R_k = \frac{E_g}{I_p}, \tag{35}$$

where $R_k$ = value of cathode resistor in ohms;
$E_g$ = desired bias voltage in volts;
$I_p$ = plate current of the tube in amperes.

It should be remembered, in the case of a tetrode or pentode amplifier tube, that the screen grid also draws current. The screen-grid current also flows through the cathode resistor. The total current flowing through this resistance for such tubes is, therefore, the sum of the plate and screen-grid currents. Formula (35) can then be altered for amplifiers utilizing screen grid tubes as follows:

$$R_k = \frac{E_g}{I_p + I_{sc}}. \tag{36}$$

where $R_k$ = value of cathode resistor in ohms;
$E_g$ = desired grid bias in volts;
$I_p$ = plate current in amperes;
$I_{sc}$ = screen grid current in amperes.

Another widely used method of obtaining bias in multistage amplifiers is the so-called **back-bias method.** The principle of back bias is identical with that of self-bias, except that a single resistor is inserted in the common cathode return lead of *all* tubes. This is shown in Fig. 160(c). The value of the back-bias resistor is obtained in the same manner as that of a self-bias resistor, except that the plate and screen-grid currents of *all* tubes being biased must be added to obtain the total current flowing through the resistor. Back-bias and self-bias methods have the advantage that variations of plate current are to some extent compensated by automatic adjustment to the bias.

The back-bias circuit shown in Fig. 160(c) is applicable when all the tubes are to receive the same bias voltage. Variations of this method are employed when it is necessary to supply different values of bias to the tube under control. One arrangement makes use of a tapped bias resistor. The proportion of the voltage drop across the back-bias resistor applied as bias will then depend upon the proportion of resistance included between the grid side of the resistor and the desired tap.

In order to prevent rapid changes in plate current from producing counteracting voltages in the grid-cathode circuit, it is customary to shunt the cathode resistor with a capacitor. The value of this capacitance for r-f amplifiers is so chosen that a sufficiently low value of reactance is offered at the lowest radio frequency within the limits of the amplifier. The r-f signal voltage is therefore offered a low-impedance path to the grid-cathode circuit. The capacitor, of course, does not affect the d-c values of bias voltage. This action is called **by-passing,** and capacitors utilized for this purpose are known as **cathode by-pass capacitors.** By-pass capacitors serve a double purpose in back-bias circuits. With no by-pass capacitor in a back-bias circuit, the r-f signal voltage would easily find its way back to the input circuit of a preceding stage through the common cathode lead. Distortion, regeneration, and possible oscillation are prevented by the use of by-pass capacitors in such circuits.

## THE A-F AMPLIFIER

Audio frequency amplifiers may be divided into three general classifications according to the method of interstage coupling utilized in each case. These classifications are resistance-coupled amplifiers, impedance-coupled amplifiers, and transformer coupled amplifiers. Each method of amplification has its advantages and disadvantages. Each of the three classifications may be further subdivided into two classes, namely, voltage amplifiers and power amplifiers.

The final stage in all classes of receiver audio amplifiers is a power

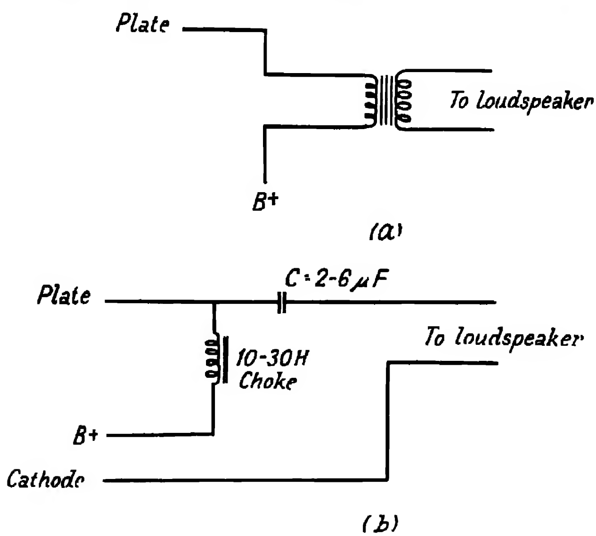



FIG. 161. Receiver output coupling systems. (a) Transformer method. (b) Choke coil method.

amplifier. It is from this stage that *power* is required to drive a loud speaker, headphones, or some other form of sound conversion device. It is essential, therefore, that a proper impedance match be established between the output impedance (plate impedance) of the power tube and the speaker or other sound conversion unit. It has been shown (Chap. X) that maximum undistorted output is obtained from a power tube when the load resistance is made approximately twice the plate resistance of the tube. In all classes of a f amplification, resistance-, impedance-, and transformer-coupled, this output-impedance match is usually most efficiently accomplished by means of transformer coupling between the plate circuit of the output tube and the loudspeaker. Such an output coupling arrangement is shown in Fig. 161(a). The proper impedance match between tube and speaker is established by applying the transformer-impedance ratio formula discussed in Chap. VI and selecting a

transformer of the proper turns ratio. Thus, if proper operation necessitates a load impedance of 10,000 ohms and it is desired to energize a speaker having a voice-coil impedance of 20 ohms, the transformer-turns ratio is found as follows:

$$\frac{Z_p}{Z_s} = (\text{T.R.})^2. \tag{37}$$

where $Z_p$ = primary reflected impedance;
$Z_s$ = secondary impedance;
T.R. = turns ratio of the transformer.

Substituting,

$$\frac{10{,}000}{20} = 500 = (\text{T.R.})^2. \tag{38}$$

Then

$$\text{T.R.} = \sqrt{500} = 22.3. \tag{39}$$

Consequently, a transformer having a primary-to-secondary turns ratio of approximately 22.3 must be employed for the above application.

Power-output stages of receivers are also often coupled to speakers by means of the choke-capacitor arrangement illustrated in Fig. 161(b). This coupling consists of an iron-core a-f choke that is placed in series with the plate of the tube and the B supply. The inductance of this choke is usually kept at a value of not less than 10 h, and offers a very low resistance to the d-c component of the plate current.

By winding the choke coil with sufficiently heavy wire, the resistance of the coil can be kept so low that the d-c voltage on the plate is comparatively unaffected by its presence. At the same time, the choke offers a high value of opposition to the a-f component of the plate current owing to its high reactance at audio frequencies. A by-pass capacitor of 2 to 4 $\mu f$ capacitance supplies a low impedance path to the speaker windings for the audio signal voltage.

In all types of audio amplifiers, all the stages preceding the power stage are voltage amplifiers. The primary function of these stages is to amplify the a-f voltage delivered by the receiver detector circuit to a value sufficiently high to swing the grid of the power amplifier tube properly. The voltage gain necessary to accomplish this swing can be represented as the ratio of the grid bias of the power tube to the a-f voltage output of the detector.

**The Resistance-coupled Amplifier.** It was shown in Chap. X that in order to obtain maximum *voltage* output from a vacuum-tube amplifier, it is essential to make the resistance of the load as high as possible. One method of doing this is to place in the plate circuit of the amplifier a resistor having a very high value of resistance. The higher the value of this resistor, the higher the signal-voltage drop across the resistor because

of the plate current flowing through it. The voltage across this resistor is then impressed across the input circuit of the following tube by means of a coupling capacitor. The capacitance of the coupling capacitor is made large enough to offer a low reactance to the a-f alternating voltage being amplified. At the same time, this capacitor prevents the high d-c plate voltage of the preceding tube from reaching the grid of the second tube. A resistor is inserted between the grid and ground of the following tube to permit d-c bias voltage to be applied to the grid of this tube. Since no direct current flows through the grid circuit, the value of this resistor has no effect on the value of the bias voltage developed across the cathode resistor. The grid resistor, or grid leak, is given a very high resistance in order that the shunting effect of the grid leak and coupling capacitor upon the plate resistor may be small. The plate resistor in a resistance-coupled amplifier is customarily called the **coupling resistor,** since it is the resistor common to both stages that couples the circuits.

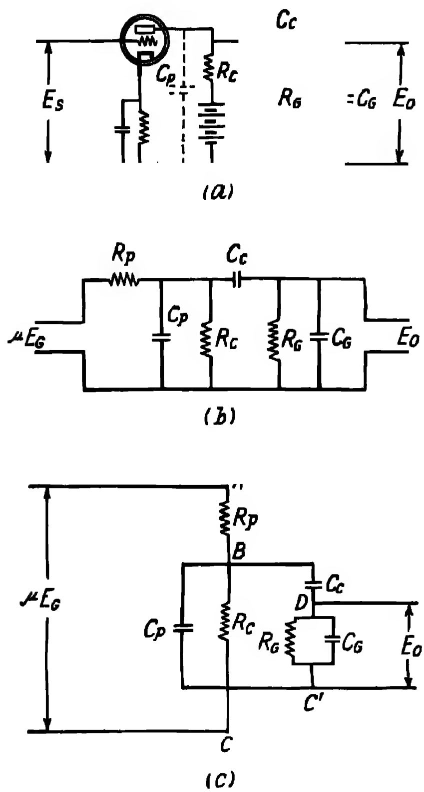



FIG. 162 (a) Triode resistance-coupled amplifier. (b) Equivalent circuit of (a) (c) Series parallel network represented by (b)

A diagram of a typical triode resistance-coupled amplifier is shown in Fig 162(a) Such an amplifier is characterized by a frequency response that is fairly constant over a wide range of frequencies but falls off at both very low and very high frequencies. The equivalent circuit for this amplifier is shown in Fig. 162(b), where

$\mu E_g$ voltage developed by tube:
$R_p$ - plate resistance of tube;
$C_p$ = plate cathode interelectrode capacitance of tube;
$R_c$ = plate coupling resistor;
$C_c$ = interstage coupling condenser;
$R_g$ -- grid resistor of following stage;

$C_g$ = grid-cathode interelectrode capacitance of succeeding tube;
$E_o$ = output voltage, or voltage developed across grid cathode of following tube.

An inspection of this circuit will reveal that it can be redrawn in the form of Fig. 162(c), which more readily lends itself to analysis.

It is at once apparent that the arrangement is a more or less series-parallel arrangement. The input voltage is applied across points $A$ and $C$, and the output voltage is taken from points $D$ and $C'$. In order to develop maximum output voltage for a given input voltage, certain steps are necessary. First, in the series circuit $ABC$, maximum voltage must be developed across points $BC$. Since it is a series circuit, the impedance of $BC$ must be kept high in comparison with $AB$. Voltage developed across $AB$ is lost to the output, since $R_p$ represents the plate resistance of the tube. The circuit of $BC$ is a parallel network. Its total, or resultant, impedance is therefore a function of the impedance of the branches. If any of these branch impedances is abnormally low, the impedance of the entire circuit $BC$ will be low, and a resultant loss of voltage will develop across it. Since the first branch consists of the interelectrode capacitance $C_p$, it is obvious that it is necessary to keep this capacitance as low as possible in order to maintain the capacitive reactance at a high value. The middle branch consists of the coupling resistor $R_c$. The highest value it is possible to maintain in this resistor depends upon practical circuit considerations, such as the permissible d-c voltage drop, and so on. The third branch of the parallel combination is in itself a series parallel circuit. Since the output voltage is taken from only *part* of this circuit, it is essential that this part $DC''$ be kept at a high value of impedance and that the reactance value of $C_c$ be kept very small. If the capacitive reactance of $C_c$ can be made small compared with the impedance $DC''$, the voltage developed across $DC''$ will become substantially equal to the voltage developed across the entire parallel network $BC$. With the coupling capacitor reactance fixed at so low a value, the impedance of the third branch of the parallel circuit $BC$ will exist almost entirely in the network $DC''$. Since $DC''$ is itself a parallel circuit, it is necessary that the impedance of both its branches be kept as high as possible, which is accomplished by keeping the value of grid resistor $R_g$ as high as is possible and consistent with practical circuit considerations. The input interelectrode capacitance of the following tube $C_g$ must be kept as small as possible in order to develop the highest possible reactance across this leg of the circuit.

For given tubes, the values $C_p$ and $C_g$ are fixed. The only variable circuit constants are, therefore, the plate-coupling resistor $R_c$, the grid resistor $R_g$, and the coupling capacitor $C_c$.

The choice of coupling resistance is dependent upon two factors: the maximum possible voltage amplification and the permissible amount of distortion. If other things remained equal, the highest voltage amplification would be obtained by using the highest possible coupling resistance for a given power supply. Normally, however, receivers have fixed plate-voltage supplies. Increasing the coupling resistance consequently decreases the plate current until a point is reached where distortion becomes serious. Beyond a certain point, very little increase in gain is obtained by increasing the coupling resistance. It is common practice to utilize in high-mu triode circuits coupling resistors that are approximately equal to the plate resistance. For triodes having lower amplification factors, the coupling resistor is commonly two to four times the value of the plate resistance. In circuits utilizing pentode tubes, the variation in voltage amplification with coupling resistance is greater than for triodes because of the larger plate resistance. However, with a fixed plate-voltage supply, the permissible d-c voltage drop across the coupling resistor becomes important. If the plate voltage falls below the screen-grid voltage, loss of amplification and distortion due to secondary emission occur. The value of plate coupling resistance is therefore limited by this factor. Coupling resistors in the order of 100,000 to 500,000 ohms are common values for pentodes.

The value of permissible grid resistance is limited by the effects of ionization in the tube. Since the grid of a resistance-coupled amplifier tube is normally run negative, positive ions released from residual gas within the tube by bombardment of electrons are attracted to the grid. If the resistance of the grid circuit is high enough, this action becomes cumulative. The control grid is therefore in danger of suddenly becoming positive, resulting in excessive plate current and possible destruction of the tube. The maximum allowable value of grid resistance depends upon the tube characteristics and the conditions under which the tube is operating. With self-bias larger values of grid resistance are permissible than with fixed bias. Values of 250,000 to 500,000 ohms are quite common for self-biased tubes.

At high audio frequencies, the reactance of the interelectrode capacitance $C_p$ and $C_g$ becomes quite low, thus effectively shunting their respective parallel networks and causing a falling-off in the gain. At intermediate frequencies, these reactances are quite high, and the resultant gain is higher. At low frequencies, the reactance of the coupling capacitance $C_c$ becomes high, and the gain again falls off. Figure 163 is a typical frequency response curve for a resistance-coupled amplifier.

**The Impedance-Coupled Amplifier.** An impedance-coupled amplifier is actually a resistance-coupled amplifier in which the coupling resistor has been replaced by an impedance, usually a high-impedance choke coil. A typical triode impedance-coupled amplifier circuit is shown in Fig. 164.

The main advantage of impedance-coupled amplifiers over resistance-coupled amplifiers is the negligible d-c voltage drop across the coupling impedance. In the impedance-coupled amplifier, this d-c voltage drop depends solely upon the d-c resistance of the choke coil, which can be brought to a negligible value by the use of fairly large wire. Impedance coupling is not very widely used owing to its greater cost.

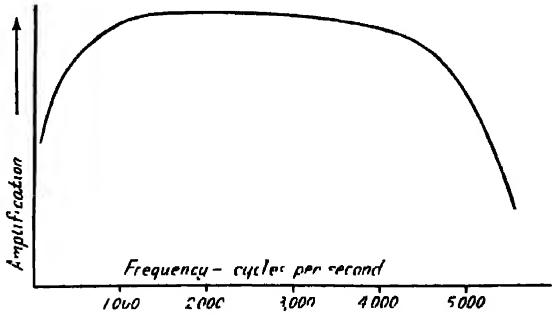



FIG. 163. Frequency response curve for pentode resistance coupled amplifiers.

Impedance coupling is characterized by high and relatively constant amplification in the middle range of frequencies. At low frequencies, the inductive reactance of the choke becomes smaller with a subsequent falling-off in amplification. At high frequencies the amplification falls off because of the shunting effect of the tube capacities as in resistance-coupled amplifiers.

The so-called double impedance coupled amplifier is one that makes

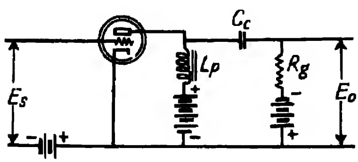



FIG. 164. Impedance coupled amplifier circuit.

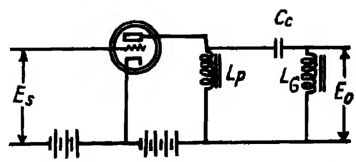



FIG. 165. Double impedance coupled amplifier circuit.

use of choke coils in both the plate circuit and the grid circuit, as shown in Fig. 165. The decrease in amplification at low frequencies can be offset in such an amplifier by choosing the values of $C_c$ and $L_g$ so that these two components form a series resonant circuit at the l-f limit of the amplifier. The rise in current through $L_g$ as resonance is approached, therefore, partly neutralizes the decrease in voltage that would normally result from the decreased reactance of $L_g$ at low frequencies. This type of coupling is relatively impractical from the economical standpoint.

**The Transformer-Coupled Amplifier.** In the transformer-coupled amplifier, the primary winding of a transformer is connected in series

with the plate circuit of the amplifier tube, and the secondary winding is connected in series with the grid circuit of the succeeding tube. The load impedance of the first tube is therefore composed of the reactance of the primary winding and the reflected impedance of the secondary load. Transformer-coupled audio amplifiers are characterized by a gradual falling off in gain at low frequencies, relatively constant amplification in the middle range of frequencies, and a sharp falling off in gain at high frequencies. The drop in amplification at low frequencies is due to the low reactance of the primary winding of the transformer at low frequencies. The more pronounced loss at high frequencies is a result of the leakage inductance and the shunting effect of the distributed capacitance of the transformer and the reflected impedance of the secondary load, including the input capacitance of the succeeding stage.

Throughout the middle range of frequencies, the effects of capacitive reactance due to distributed capacitance and input capacitance of the following tube and the inductive reactance of the primary tend to neutralize each other. As the frequency is increased, the inductive reactance *increases* and the capacitive reactance *decreases*. Conversely, as the frequency is decreased, the inductive reactance *decreases* and the capacitive reactance *increases*; that is inductive reactance and capacitive reactance vary inversely with each other for any change in frequency. By far the largest portion of reactance, both inductive and capacitive, is in the transformer itself. Consequently, the frequency limits of a transformer-coupled amplifier are dependent to a large extent upon the transformer design. Modern well-designed audio transformers readily cover the voice frequency range with comparatively little variation in amplification. Where a very wide range of frequencies is to be transmitted, however, resistance coupling is preferable.

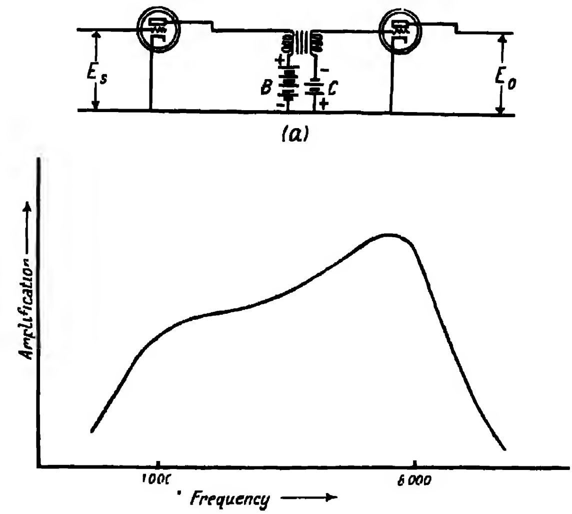
One of the main advantages of transformer coupling is the additional gain obtainable by increasing the turns ratio. Losses attendant upon increased turns ratio, however, sometimes counterbalance the advantages so obtained. Increasing the turns ratio limits the permissible amount of inductance in the primary. In addition, the shunting effect of input capacitance becomes more pronounced as the result of the greater impedance ratio. Increasing the turns ratio of a transformer, therefore, decreases the usable portion of its frequency response curve. A typical frequency response curve for a transformer-coupled amplifier is shown in Fig. 166(b).

On the whole, transformer coupled amplifiers have been largely superseded by resistance-coupled amplifiers. The latter are much more economical, and with the high-mu tubes now available, satisfactory gain per stage is easily obtained without the use of step-up transformers. Transformer coupling, however, is to be preferred for certain applications, such as push-pull amplifiers (see page 298). Modern receivers invariably

utilize transformer coupling between output stage and loudspeaker. Figure 166(a) illustrates a typical transformer-coupled amplifier circuit.

## THE TUNED R-F RECEIVER

A tuned r-f receiver is composed of three sections, namely an r-f amplifier, a detector, and an audio amplifier. The minute signal voltage developed in the receiving antenna system is amplified by successive




(b)

FIG. 166 (a) Transformer-coupled amplifier circuit (b) Frequency response curve for typical transformer coupled amplifier.

stages of r-f voltage amplifiers. This amplified voltage is then impressed across the input circuit of the detector stage. The a f component of the detector output is impressed across the input circuit of the a-f amplifier. This amplifier may also consist of several stages, the last of which is a power amplifier. The a-f voltage impressed across the first a-f amplifier stage is amplified until it is large enough to swing the grid of the power-amplifier tube efficiently. The power-amplifier tube is, in turn, coupled to a sound conversion device such as a loudspeaker. This latter unit converts the a-f power in the form of electric energy into power in the form of sound waves.

It is standard practice to begin the design of such a receiver with the output and work backward to the input stage. To fulfill certain functions in a specific application, a receiver must meet certain requirements. These requirements are referred to in terms of power output, sensitivity, selectivity, and fidelity.

The **power-output rating** of a receiver is the maximum *undistorted* power output that the receiver can handle. Since all vacuum-tube circuits produce *some* distortion, the term "undistorted power output" must be qualified. Undistorted output has been standardized by the radio-engineering profession to mean an output that does not contain enough distortion to become objectionable to the human ear. This limiting amount of distortion has been standardized by the Institute of Radio Engineers as 10 per cent. Thus, the maximum undistorted power output of a receiver is defined as the maximum power output that can be obtained from a receiver when the output voltage does not contain more than 10 per cent of harmonic distortion This critical output level is often called the **overload level** of a receiver.

The **sensitivity** of a receiver is a measure of its ability to receive weak signals. The greater the sensitivity, the weaker the signals that will be received. Sensitivity is measured quantitatively in terms of the input-signal voltage required to produce a certain output The conditions under which sensitivity is measured have been standardized by the profession. All receiver sensitivities are calibrated for standard test outputs. Standard test output is 0.05 w of a-f power in a noninductive resistor connected across the output terminals of the receiver. The resistance used should have whatever value is necessary to obtain maximum undistorted power output for the type of output tube under consideration.

The **selectivity** of a receiver is a characteristic not easily measured. It is usually expressed in one or more graphs depicting the frequency response at a response test frequency and the response at frequencies higher and lower than resonance. Selectivity is usually measured under the same conditions as sensitivity, that is, the receiver output is maintained at test level (0.05 w). The input signal frequency (supplied by a laboratory signal generator) is then varied from a point 100 kc below resonance, through the resonant frequency, to a point 100 kc above resonance. The frequency changes are made in steps of 10 kc or less. After each change the input voltage is increased (or decreased) until the power output is 0.05 w. The various signal voltages necessary to achieve this result throughout the test spectrum are then plotted in a graph. The resulting curve is a fairly accurate representation of the selectivity of the receiver. Often such a frequency response curve is made by plotting decibels against frequency, since decibels are units more indicative of the usability of the receiver. The use of decibels is discussed in detail on page 392.

The **fidelity** of a receiver is a measure of the degree to which it accurately reproduces in its output the signal that is impressed upon it. Fidelity is measured by applying a 30 per cent modulated signal to the receiver by means of a signal generator The modulation frequency is varied from 40 to 10,000 c in convenient steps. Output

readings are taken at each step and plotted against modulating audio frequencies.

A typical tuned r-f receiver circuit is illustrated in Fig. 167. Maximum selectivity is derived from such a circuit by tuning each of the r-f amplifier stages in addition to the detector stage, from which the receiver obtains its name.

In the early days of radio, the tuned r-f receiver was very widely used. One of the chief disadvantages of this receiver was the necessity of employing neutralizing circuits to prevent self-oscillation in the r-f

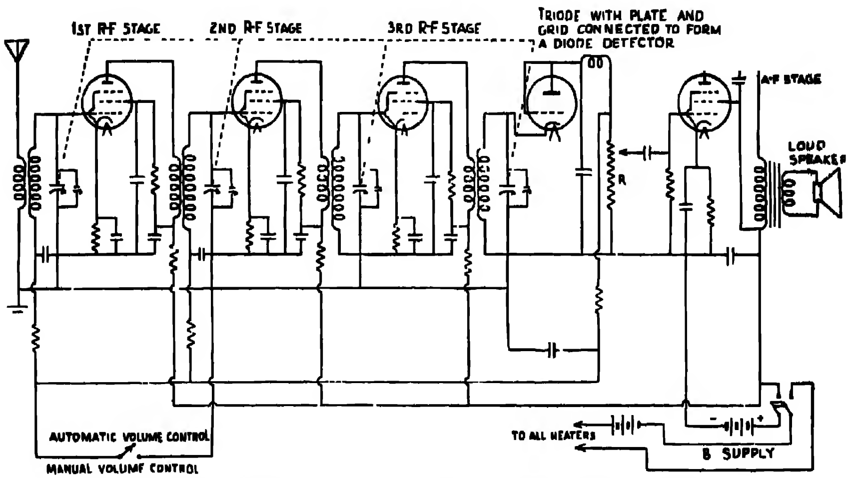



FIG. 167 Tuned r-f receiver. (*Courtesy of RCA Mfg. Co., Inc.*)

amplifier. The introduction of the screen grid tube did away with this difficulty and at the same time permitted greater amplification to be obtained in the r-f stages. The pentode tube also permitted further improvement in this direction.

As soon as certain patent litigation involving the superheterodyne circuit was eliminated, manufacturers decreased their outputs of tuned r-f receivers in favor of the superheterodyne. Because of the superior sensitivity and selectivity of the superheterodyne receiving circuit, the use of the tuned r-f receivers is comparatively limited today.

Early tuned r-f receivers utilized individual tuning condensers for each r-f stage and the detector. Proper operation of the receiver involved tuning individually each stage to resonance, a system that necessitated manipulation of numerous controls each time the receiver was tuned to a different station. This situation was improved by later designs utilizing gang control of the tuning capacitors. Modern receivers utilize ganged capacitors in a single unit having the rotors of each capacitor section mounted upon a common shaft.

In order that a tuned r-f receiver utilizing ganged capacitors may function at peak efficiency, all the tuned circuits must track, that is, all the tuned circuits must resonate at the same frequency for any given setting of the dial. Standard practice is to use identical transformers and capacitors in such receivers for this reason. Nevertheless, owing to minor circuit differences in wiring, shielding, and so on, it is usually necessary to provide a supplementary means of bringing the tuned circuits into alignment. This is accomplished by the insertion of a small semifixed capacitor, called a **trimmer,** in parallel with each main tuning condenser. Minor discrepancies in the resonant frequencies of the various circuits are then compensated for by variation of the trimmer capacitance. Since the trimmer is shunted across the main resonant circuit, any variation of it will vary the resonant frequency of this circuit. Trimmers may be mica- or air-insulated, but are more often the former. They are customarily manufactured so that adjustment of their capacity is accomplished by means of a special tool, usually an insulated screw driver and, once adjusted, will remain fixed at this value of capacitance.

## *THE SUPERHETERODYNE RECEIVER*

**Principle of Operation.** The popular superheterodyne receiving circuit depends for its operation upon the phenomenon of heterodyning that was discussed in Chap. XI In general, the system operates as follows: The incoming radio signal is heterodyned with a local oscillator in the receiver whose frequency differs slightly from the signal frequency. Both these signals are fed into a common stage, variously called the **converter, mixer,** or **first detector** stage The output circuit of this stage and the tuned circuits of the succeeding r-f stages are tuned to resonate at the *beat* frequency In order not to confuse the r-f stages preceding the mixer stage with those following it, the latter are called **intermediate-frequency** (or **i-f**) stages. The i-f amplifier feeds into a conventional detector circuit followed by an audio amplifier.

One of the chief merits of this system is the high degree of selectivity obtainable. Much better discrimination against unwanted interfering signals is possible than with any other receiving circuit. For example, assume a desired signal of 900 kc is to be received in the presence of interference from a 910 kc signal. The two signals differ in frequency by approximately 1 per cent. If the local receiver oscillator is adjusted to 950 kc, the beat, or intermediate frequency, produced with the desired 900-kc signal is 50 kc. The undesired interfering signal of 910 kc will produce a beat frequency of 40 kc with the local oscillator. The two original signals, which differed in frequency by only 1 per cent, have been converted to 50-kc and 40 kc signals, which differ in frequency by 20 per cent. When both of the latter signals are impressed on the i-f amplifier

(which is tuned to 50 kc), the 40-kc signal is easily rejected. When the intermediate amplifier consists of more than one stage, the superheterodyne receiver becomes extremely selective.

In actual commercial receivers, intermediate frequencies as low as 50 kc are seldom encountered, because of other circuit complications which are discussed later in this section.

The tuning condenser in the tank circuit of the local oscillator in a superheterodyne receiver is ganged to the tuning condenser of the mixer input circuit. The constants of each circuit are so selected that the oscillator frequency differs from the input circuit frequency by a fixed amount, regardless of the frequency to which the receiver is tuned. Efficient superheterodyne circuits also customarily utilize one or two r-f stages ahead of the mixer stage. These are also ganged to the mixer input and oscillator stages. The r-f and mixer input circuits, of course, tune to the same frequency. By virtue of the fact that this arrangement fixes the beat frequency produced, regardless of the frequency of the incoming signal, it is possible to construct the intermediate amplifier circuits to tune to a fixed frequency. This type of construction permits the use of tuned circuits in the amplifier having exceedingly high circuit $Q$ with all the attendant advantages. As a result, it is possible to obtain as much amplification with a single stage of i f amplification as with three or four stages of variable tuned amplifiers.

In a variable-tuned amplifier, such as used in the r f sections of tuned r f receivers, the tuned circuit $Q$ is continually varied as the circuit is tuned. Consequently, it is possible to obtain maximum $Q$ at only one frequency within the range of the circuit. At all other frequencies to which the circuit will tune, there will be lesser values of $Q$. In other words, both the selectivity and the amplification vary at different parts of the band to which such circuits tune This difficulty is largely corrected in the superheterodyne receiver, since the i f amplification takes place with constant amplification and with the same degree of selectivity, because the circuit $Q$ is always fixed. The over all characteristic of the receiver is limited in this respect by the variable tuned stages in the r-f amplifier and the mixer circuit. Nevertheless, since by far the greater part of the selectivity and amplification is obtained in the i f amplifier, the over-all characteristic of the superheterodyne is far superior to the tuned r-f receiver.

The design of circuits that provide fairly constant tracking of oscillator and input stages sometimes presents a problem in superheterodynes, especially on the higher frequencies where oscillator stability is not good. Proper tracking is sometimes accomplished by having the rotor plates of the oscillator tuning capacitor displaced in rotation on the shaft with respect to the rotor of the r f capacitors. A variation of this idea is to have the oscillator capacitor rotor smaller in size than the r-f capacitor

rotors. Such systems are fairly effective on the lower frequencies but present difficulties on the high frequencies. An improved plan utilizes a coil having slightly lower inductance for the oscillator. In such circuits the final adjustment is accomplished by small shunt and series capacitors (called trimmers and padders respectively) connected to the parallel resonant circuits. By meticulous adjustment of trimmers and padders it is possible to make such circuits differ by the exact required intermediate frequency on at least three points of the dial. The fluctuation over the remaining intervening sections of the frequency range is comparatively small and can usually be neglected. A commercial fixed frequency superheterodyne receiver is shown in Fig. 168.

**Problems Peculiar to the Superheterodyne.** One of the chief troubles to which the superheterodyne receiver is subject is **image frequency response,** a result of the fact that at any dial setting there are two possible signal frequencies that will produce the necessary intermediate frequency Thus, if a given receiver utilizes an intermediate frequency of 200 kc, the local oscillator will be tuned to 1,200 kc when receiving a desired 1,000-kc signal. If the circuits preceding the mixer stage are insufficiently selective, a powerful local station operating on 1,400 kc can also supply sufficient signal to the mixer circuit grid to produce another 200 kc beat frequency. This signal would receive the same amplification in the i-f amplifier that the desired signal receives, with resulting severe interference. The only remedy for image frequency response is to prevent the undesired signal from reaching the mixer tube. There are two ways of accomplishing this. The first is to use sufficient r f amplification ahead of the mixer so that the circuit is selective enough to reject the undesired signal. The second way is to use an intermediate frequency high enough to remove the image response frequency a considerable amount from the resonant frequency of the input circuit. If this separation is great enough, a comparatively small amount of selectivity in the input circuits will suffice to reject the unwanted signal.

In considering the latter remedy, a happy medium must be attained in choosing the proper intermediate frequency. The frequency must be high enough to provide sufficient image-frequency rejection and at the same time low enough to provide good i f selectivity, as discussed on page 292.

Another method of suppressing image frequencies is by the use of series or parallel rejector circuits (wave traps) appropriately connected to the input circuit of the first r f stage. Such circuits, of course, must be variable-tuned, must be ganged to r-f and oscillator circuits, and must be properly adjusted to track with r-f circuits in order to maintain a fixed frequency difference. Several manufacturers have marketed special circuits to attain this end.

Another type of interference often experienced in superheterodynes is

that due to harmonics generated by the local oscillator. Thus, if the oscillator produces a strong second harmonic, this harmonic may beat with an undesired signal frequency considerably higher than the desired signal frequency to produce a heterodyne equal to the intermediate frequency. One remedy for this condition is the same as for image frequency response, namely, to increase the preselection before the mixer circuit. The best remedy, obviously, is to eliminate the harmonics from the local oscillator output. This can sometimes be accomplished by decreasing the oscillator plate voltage, thereby decreasing the oscillator

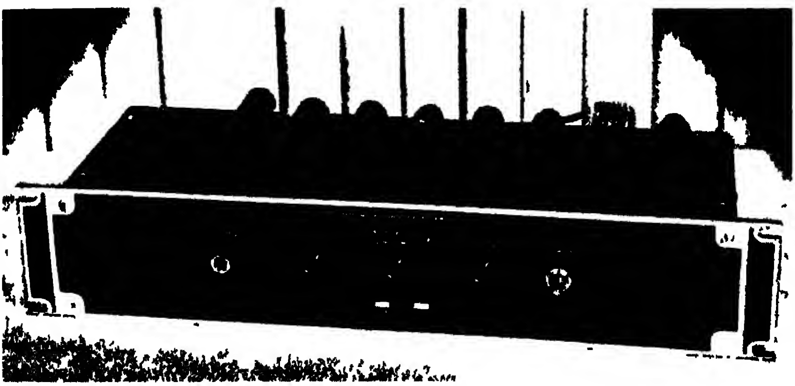


FIG. 168. Commercial radio telegraph fixed-frequency receiver. The heterodyne oscillator in such receivers is crystal controlled and operates at only one frequency. All *LC* circuits are nontunable. (*Courtesy of Communications Co., Inc.*)

output on all frequencies. Since the harmonics are always a comparatively small percentage of the amplitude of the fundamental, this usually brings the harmonic output down to a negligible value. The necessary fundamental oscillator output is small anyway, and there is no need to exact any appreciable amount of power from this oscillator. Another scheme is to insert a fairly large resistance in the oscillator plate circuit. This resistance will tend to straighten the oscillator tube characteristic curve with a resultant decrease in harmonic content.

Another fault sometimes encountered with superheterodynes having insufficient preselection is **double-spot tuning,** which consists of receiving the same station at two different places on the dial. Double spot reception occurs when the local oscillator is tuned intermediate frequency above or below the frequency of a given station. Thus, if a receiver having an intermediate frequency of 200 kc is tuned to receive a 1,000 kc station, the local oscillator will be tuned to 1,200 kc, producing a 200-kc beat frequency. When the receiver is tuned to 600 kc the oscillator will be tuned to 800 kc. If the input circuit is insufficiently selective, the 1,000-kc signal will get through to the mixer input circuit to beat with the 800-kc

oscillator and again produce a 200-kc beat frequency. Consequently, the same station will be heard at two places on the dial, that is, at 1.000 kc and at 600 kc. All the remedies that apply to image-frequency suppression apply equally well to suppression of double-spot tuning.

The functions of mixer and local oscillator are often combined in a single tube. A number of multigrid vacuum tubes, such as the 2A7, 6SA7, 6A8, 6K8, and 6L7, have been especially developed for this purpose. In such circuits, two of the grids act as plate and grid of a triode oscillator in conjunction with the common cathode. The remaining grids and the plate perform the function of mixer tube. The voltage developed by the oscillator acts to modulate the electron stream flowing to the mixer section of the tube. This system has many of the advantages of the electron-coupled oscillator discussed in Chap. XI.

## GENERAL RECEIVER CONSIDERATIONS

**Volume-control Methods.** In the early days of battery radio receivers, the generally accepted method of controlling volume was by varying the filament emission of the tubes. With the advent of indirect-heater type of tube and improved circuit design, this system was no longer possible owing to the lag in cathode emission with a change in heater current. This system had the additional disadvantage of varying the effective operating point of a tube along its characteristic curve. For a time thereafter, the approved method was varying audio amplifier gain or shunting the loudspeaker. It also became standard practice for a period to control volume by varying the gain of the r f and i-f circuits. Some manufacturers control volume by shunting the antenna input circuit, others make use of various combinations of two and sometimes three of the above methods.

A number of typical modern volume-control circuits are shown in Fig. 169. Figure 169(a) illustrates an *antenna shunt* type of volume control. This consists of a potentiometer which is effectively shunted across the input of the receiver. The fixed contacts are connected to antenna and ground respectively. The primary of the input r f transformer is connected to ground and to the moving contact of the potentio meter. When the control is at maximum position, the total resistance of the potentiometer is in shunt with both the antenna circuit and the primary coil. Current due to a signal voltage on the antenna flows to ground through the potentiometer resistance, causing a voltage drop across it. The position of the movable contact determines the portion of this voltage that is impressed across the r-f transformer primary. The total resistance of the control must be of such a value that at full-volume position very little current flows through the potentiometer branch of the parallel circuit. In practice, the control is usually made four or five

times the value of the primary impedance. Higher values are not practical because of the shunting action of the antenna impedance.

Figure 169(b) illustrates a method of controlling volume by varying r f bias. Increasing the bias of a tube lowers the mutual conductance and decreases the gain of the stage. Several stages of amplification may be controlled with a single potentiometer utilizing this method. Bias control cannot be obtained with tubes having a sharp cutoff.

A combination antenna and bias control system is shown in Fig. 169(c). Two distinct actions are combined in this circuit. One is the

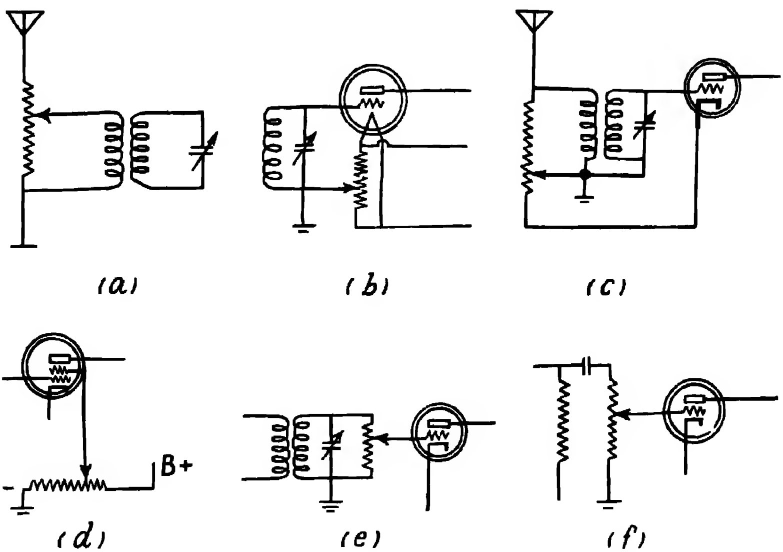



FIG. 169 Volume control circuits (a) Antenna shunt. (b) R F bias (c) Antenna bias (d) Screen grid control (e) R F shunt (f) A F shunt

control of volume by increasing the bias on the controlled tubes, the other is the shorting out of the input signal at the antenna. Often duo potentiometers are utilized for this double function. Such potentiometers are really two separate units mounted on a common shaft. This combination circuit is used where the straight antenna shunt type of circuit would not give full attenuation of powerful signals.

Figure 169(d) is an illustration of a commonly used screen voltage control. In most respects, the action in this arrangement is similar to that obtained with bias control. Varying the screen voltage varies the mutual conductance and hence the gain.

In Fig. 169(e) is shown the r-f shunt method of control. The full resistance of the potentiometer in this case is shunted across the secondary of a tuned r-f transformer. Since the grid of the tube is connected to the

movable arm of the control, any variation of this arm varies the amount of r-f voltage impressed on the grid of the tube.

As in other shunt type circuits, the resistance of the potentiometer should not be of a value low enough to present too great a load to the r-f voltage developed in the transformer secondary. The lowest permissible value is approximately 100,000 ohms. Values of 250,000 and 500,000 ohms are quite common for this type circuit.

Figure 169(f) shows another form of shunt control used in audio-amplifier circuits This circuit is for a resistance coupled amplifier. The potentiometer is actually part of the plate load of the preceding tube. as discussed in an earlier part of this chapter. The resistance of the control is therefore determined by the required plate load of the preceding tube in conjunction with the plate resistor of this tube and the coupling condenser.

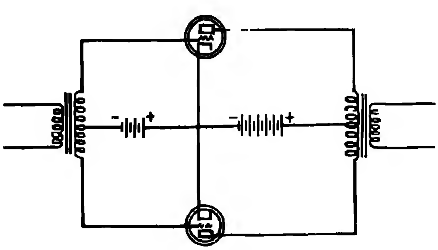


FIG. 170 Transformer coupled push pull amplifier

This shunt type of volume control is also used frequently in transformer coupled audio circuits. The circuit resembles that of Fig 169(e) The resistance of the control is determined primarily by the plate load of the preceding tube. The impedance ratio of the transformer must therefore be considered in calculating the potentiometer resistance

**Push-pull Amplifiers.** An advantageous system of audio power amplification often used for the power output stage of both tuned r-f and superheterodyne receivers is the push pull amplifier A diagram of the fundamental push pull circuit is shown in Fig 170 By means of the center tap on the secondary of the input transformer, the grids of the two tubes are excited with equal voltage 180° out of phase. The outputs of the two tubes are combined by means of the center-tapped output transformer

The tubes used in push pull amplifiers are customarily operated either class B, class A, or class AB. When operating class B, each tube is operated with a negative bias such that the plate current is reduced to approximately zero when no signal voltage is applied and plate current flows only during *positive* half cycles of the signal voltage (see Chap. X). It is apparent, therefore, that only one tube operates at a time. Since the grids are excited 180° out of phase, the first tube will operate during the first alternation of the cycle, the second tube during the second alternation of the cycle, and so on. Because of the center-tapped output transformer arrangement, the outputs of the tubes are added to each other in the proper phase relation. The signal in the secondary of the

output transformer is therefore a complete cycle of the amplified input signal.

Even with the high negative bias necessary to operate the tube at zero signal cutoff, the grid sometimes goes positive at the peaks of strong input signals. When this happens, the grid draws current. The preceding amplifier stage must therefore be designed to deliver a certain amount of power to make up the grid circuit losses in the push-pull stage. Such a preceding stage is called a **driver stage** and usually consists of a single-tube power amplifier operating class A.

Vacuum tubes have been designed especially for push-pull amplifier service. Such tubes permit cutoff to be obtained with zero bias. Thus, the tube draws grid current throughout the entire positive alternation of the input cycle, necessitating the use of a driver stage of considerable output.

The class B push-pull system offers considerable advantage when the input signal voltage is large. On weak signals, appreciable distortion occurs because of the curvature of the tube's characteristic curve.

In the class A push pull system, *both* tubes operate throughout the entire input cycle. Since the tube grids are never allowed to go positive, no current is drawn (power consumed) in the grid circuit, and, consequently, a driver stage is not needed. Furthermore, since the direct currents in the two halves of the output transformer primary windings magnetize the core in opposite directions, there is no d c saturation in this core. This condition permits the use of smaller output transformers in class A push-pull systems than in single-tube amplifiers. The class A push-pull amplifier, however, has a tendency to overload easily when the signal strength is large

The class AB push pull amplifier combines the advantages of the class A and class B types. In a class AB (sometimes called class A′) amplifier, the tubes operate class A for weak signals and class B when the input signal voltage is large. Although this system also requires a preceding driver stage, the power required from the driver is not so great as for class B operation. In general, all the considerations applying to a class B push pull amplifier apply also to a class AB push-pull system, although not to as great an extent. The input and output transformers need not be as large as those used for class B, but must conform more closely to class B requirements than to class A.

The advantages of push pull amplification compared with single-tube amplification are manifold. In the class A and class AB types, there is little or no current of signal frequency flowing through the plate-supply source. This eliminates the need for by-passing and additional filter circuits in this stage and obviates the necessity for cathode biasing resistor by-pass condensers. Even when there is a plate impedance common to this and other stages in the audio amplifier, there is no regeneration.

In addition, there is less distortion for the same power output per tube in a push-pull amplifier than in an equivalent type of single-tube amplifier. This is due to the cancellation of all even-order harmonics in the output circuit. Since by far the greatest amount of distortion in an amplifier is caused by the presence of second harmonics, excellent improvements may be obtained using push-pull amplifiers.

Alternating current hum components present in the source of plate power in conventional amplifiers usually appear in the output unless the plate-supply source is exceptionally well filtered. In push pull amplifiers, the hum currents flowing in the two halves of the output-transformer

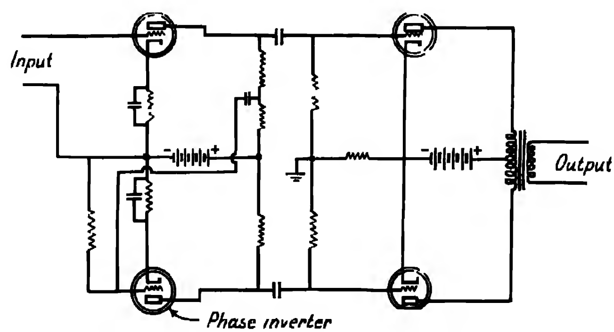



FIG. 171 Push pull resistance coupled amplifier employing phase inversion

primary balance each other out, thus reducing the amount of plate-supply filtering necessary.

Push pull output stages can also be utilized in resistance coupled amplifiers. Such circuits require the use of an additional tube but eliminate the input transformer. A typical fundamental circuit is shown in Fig. 171. A preamplifier tube supplies excitation to one of the push pull tubes by resistance coupling in the usual manner. A portion of the output voltage developed in the plate resistor of this preamplifier tube is fed to the input circuit of the additional tube, which is called the **phase-inverter tube.** The output of the phase inverter supplies excitation to the remaining push pull tube through a resistance coupling network. Since the output of any amplifier tube is 180° out of phase with its input, the excitation voltages applied to the grids of the push pull tubes are also 180° out of phase. Care must be exercized in adjusting the tap on the plate resistor of the preamplifier tube in order that the voltages supplied to the push-pull grids are *equal* in amplitude

**Feedback Amplifiers.** An ideal amplifier is one that produces an output that exactly duplicates the input in all respects, except, of course, amplitude. Unfortunately, such performance is impossible of attainment in practical amplifiers. The ordinary amplifier circuit, regardless of the type, is productive of considerable distortion of the wave form.

The types of distortion encountered in a-f amplifiers are *amplitude distortion* (often called "nonlinear distortion"), *frequency distortion*, and *phase distortion*. Amplitude distortion is the result of operating a tube over a nonlinear portion of its characteristic. Such operation results in the production of frequencies in the amplifier output that are not in the input. The most troublesome of these distortion frequencies are harmonics of the input frequencies (notably the second harmonic) and beat frequencies produced by heterodyning of the signal components. The elimination of amplitude distortion in an amplifier is a matter of proper circuit design. Anything that ensures operation of the tube only over the linear portion of its characteristic or reduces the curvature of the

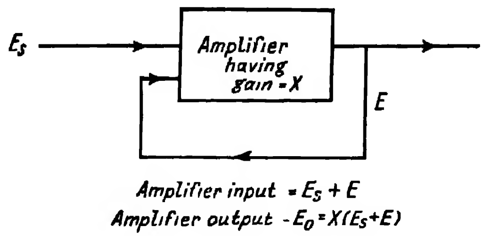



FIG. 172. Inverse feedback amplifier in block diagram.

tube characteristic will reduce amplitude distortion. This has been discussed in Chap. X and will not be treated further here.

Frequency distortion is caused by the unequal amplification of different frequencies of signal voltage. The effects of frequency distortion are more noticeable when the amplification per stage is increased. Usually, the higher signal frequencies are discriminated against and receive less amplification than the lower frequencies but where high fidelity is important, it is generally necessary to sacrifice some amplification.

Phase distortion is caused by a disturbance of the phase relations between different frequency components of the signal voltage as the signal is amplified. This disturbance is equivalent to a process in which different frequencies are transmitted through the amplifier with different velocities. Consequently, all the frequencies do not arrive at the output at the same time. The subsequent relative phase relations of the different frequency components of the output voltage are different from those of the input voltage. Considerable phase distortion can be tolerated before the effect is noticeable to the human ear. For this reason, phase distortion is not so important a factor as are the other forms of distortion.

When it is desired to achieve high fidelity in a radiotelephone receiver, it is sometimes efficacious to introduce regeneration deliberately in the amplifier in such a way as to reduce the amplifier gain. Such feedback amplifiers are commonly called **inverse-feedback amplifiers.** A block diagram of an inverse-feedback amplifier is shown in Fig. 172. A portion

$E$ of the output voltage is introduced into the input in such a way that it subtracts from the original input voltage $E_s$. Such amplifiers greatly reduce frequency and phase distortion. In addition, they act to stabilize the amplifier, making the amplification substantially independent of tube constants and electrode voltage variation. Since the regenerative voltage is fed back *inversely*, it acts to *decrease* the effective gain of the amplifier. Thus, when feedback is not present, frequencies that are normally discriminated against in the amplifier appear at lower magnitudes in the output than do lower frequencies. Therefore, the amplitude of the feedback voltage is *less* at these frequencies and, consequently, effects a *small decrease* in gain of the amplifier. The intermediate frequencies, which receive considerably greater original amplification, appear at greater magnitudes in the output. Therefore, the amplitude of feedback voltage at these intermediate frequencies is correspondingly greater, effecting a resultant *greater decrease* in gain of the amplifier. When properly designed and adjusted, inverse feedback amplifiers can be operated to deliver substantially uniform amplification at all frequencies within the range of the amplifier.

**Tone Control.** Theoretically, a receiver tuned to a broadcast station should be operated with an output such that the listener hears the program with the same volume as he would hear it if he were seated in the hall where it is originating. Actually, however, it is usually not practical to operate a radio receiver at this output. Despite the fact that a high-grade high-fidelity receiver is being employed, which transmits all passages of a typical musical program with equal fidelity, the true fidelity of the reproduction will be lost upon the listener if the receiver is operated at low volume. This is due to a peculiar characteristic of the human ear. The sensitivity of the human ear is known to vary with the *volume* of sound as well as with the frequency. It has long been known that the human ear does not have a linear frequency response characteristic. At the very low and very high frequencies, the sensitivity of the ear is much less than it is throughout the middle range of the a-f spectrum. In addition, when the volume of a sound is decreased, the sensitivity of the ear to the l-f notes decreases much more than does the sensitivity to the h-f and i-f notes. As a result, when the volume of a receiver is turned down, the net result on the human ear is an effective decrease in l-f notes. If the receiver is turned down very low, the l-f notes seem to disappear altogether.

In order to compensate for this failing, many broadcast receivers are equipped with some system of tone control. The receptivity of the human ear to high and low frequencies at different volumes appears to be a matter of relativity. Thus, it has been found that *decreasing* the amplitude of h-f tones has the same effect on the ear as *increasing* the l-f tones. In view of this peculiar physiological reaction of the ear, most tone control

circuits employed in receivers function to decrease the h-f output. This is usually accomplished by some type of variable circuit shunted across the output portion of the audio amplifier. Such circuits utilize a capacitance for the major component. The value of this capacitance is such that the capacitive reactance is small at high frequencies, and the h-f notes are therefore shunted out of the circuit. The reactance at low frequencies is relatively high, allowing these frequencies to be transmitted with negligible attenuation.

A number of typical tone-control circuits are shown in Fig. 173. One of the simplest types is shown at (a). This consists simply of a capacitor in series with a variable resistance shunted across the grid, or input, circuit of an audio-amplifier stage. When the resistance of the control

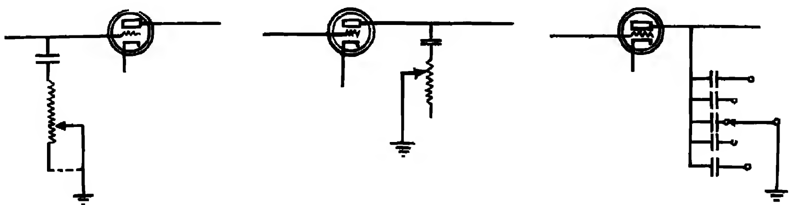



FIG. 173. Typical tone control circuits. (a) Grid-circuit control (b) Plate-circuit control (c) Plate-circuit control

is zero, the higher frequencies of the signal are by-passed and do not appear in the output. As the resistance of the control is increased, the impedance of the entire circuit is increased with resulting diminution in by-pass action. The same type of potentiometers may be used for this service as is used in volume-control circuits.

Figure 173(b) illustrates a plate circuit type of control. This type functions in exactly the same manner as the grid type of tone control previously discussed. The main difference between the plate and grid types of tone control is in the impedance of the circuit that is being shunted. Thus, in the grid type when used with a high impedance input circuit, the control circuit is shunting a grid circuit having an impedance of 100,000 ohms or more. In the plate type of control, the circuit being shunted has an impedance varying from 2,000 to 20,000 ohms. Consequently, in a grid circuit, a small capacitor and a large value of resistance must be used. In a plate circuit, a larger capacitor and a lower value of resistance are required to give the same amount of tone control.

In Fig. 173(c) another type of tone control is shown, which consists simply in shunting any of a number of capacitors of different capacity across the plate circuit by means of a switching arrangement.

Some manufacturers combine the functions of volume control and tone control in one control. Such systems are commonly called **acoustically**

**compensated volume controls.** The circuit arrangement is such that as the volume of the receiver is decreased, the l-f response is increased (h-f response decreased). Some systems utilize two potentiometers mounted on a common shaft for this purpose, and the units are arranged so that, as the resistance of one potentiometer is increased, the resistance of the other is decreased. Many manufacturers, however, utilize a single potentiometer for these combined functions. In this arrangement one side of the potentiometer and the variable tap are utilized for volume-control purposes; the other side of the control and the variable tap are utilized for the tone-control circuit. Thus, as the volume is decreased by moving the variable tap *away* from the volume side of the unit, thereby increasing the resistance of the volume control, the resistance in

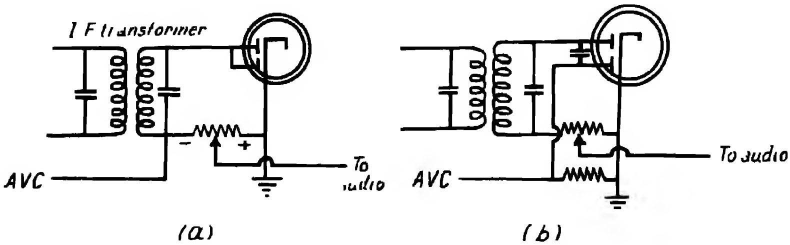



FIG. 174 Diode avc systems.

the tone-control side of the circuit is decreased thereby increasing the h-f by pass action.

**Automatic Volume Control.** Automatic volume control circuits (abbreviated avc) in a receiver act to control continuously and to adjust automatically the sensitivity of the receiver. The signal input to the audio amplifier, therefore remains essentially constant within certain limits over a wide range of received signal amplitudes. This is accomplished by biasing the grids of the r-f and i-f tubes with a negative voltage obtained by rectifying the signal carrier. Any increase in signal increases the negative bias and decreases the amplification of these stages. Any decrease in signal correspondingly decreases the negative bias, thereby increasing the amplification.

Some systems of automatic volume control utilize a portion of the detector output for bias purposes, as in the circuit of Fig. 174(a). In some other systems, which utilize double diode detection, one diode section is used for detection while the other develops the avc bias voltage, as shown in Fig 174(b). There are a great many different methods of obtaining automatic volume control in a receiver, but space limitations prevent the discussion of all the types here.

One method of obtaining automatic volume control that is used in the better receivers utilizes a separate amplifier and a separate rectifier to

obtain improved avc action. Such systems are called **amplified automatic volume control** systems.

One of the disadvantages of an avc system is the fact that it begins to operate as soon as any signal reaches the avc tube. A very weak signal, therefore, does not receive the full amplification that the receiver is capable of supplying. To overcome this limitation, some receivers incorporate circuits that prevent the avc circuit from functioning until the incoming signal exceeds a certain level. Such systems are termed **delayed automatic volume control** circuits.

In general, delayed automatic volume control is accomplished by applying a small negative voltage to the plate of the avc rectifier, as shown in Fig. 175. No current will then flow through the diode circuit until the incoming signal voltage exceeds the amount of this negative voltage, thus permitting the plate to become positive. Although a battery is utilized to supply the negative voltage in the diagram, it is customary to utilize a portion of the plate-supply bleeder circuit for this purpose.

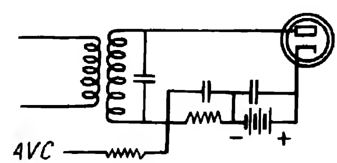


FIG. 175. Delayed avc system.

Efficient avc circuits greatly improve the performance of receivers. They eliminate the need for frequent manipulation of manual volume controls when tuning from comparatively weak to strong stations. Thus, a receiver with a good avc system can be tuned from one end of its frequency range to the other, and most of the stations will come in with substantially the same volume. The avc system is not a cure-all for insensitive receivers, however, for there will always be some signals that are so weak that they will not be amplified sufficiently to produce the desired volume. The desired volume level is determined by the setting of the manual volume control of the receiver. In sets having avc systems, the manual volume control is usually located in the a f amplifier and does not form a part of the avc system.

**Interstation Noise Suppression.** Since the normal action of an avc system is to *decrease* the receiver sensitivity on strong signals and to *increase* the sensitivity on weak signals, the sensitivity is at maximum when the receiver is tuned off resonance, that is, between stations. Any static or man-made electrical disturbances are amplified by the full amplifying power of the receiver. This interstation noise can become very troublesome and irritating when a sensitive receiver equipped with automatic volume control is operated in a noisy location.

A number of systems have been developed to overcome this difficulty, which have been variously termed "silent tuning systems," "automatic sensitivity circuits," "interstation noise suppression circuits," and "quenched," or "quiet automatic volume control circuits" (qavc). In

general, noise-suppression circuits utilize a separate tube biased by the avc voltage. A high value of resistance is common to the plate circuit of this tube and usually to one of the audio-amplifier tubes. Between stations, when no signals are tuned in, there is no avc voltage developed and hence no bias applied to the silencer tube. Plate current flows through the tube causing a voltage drop across the resistor. This voltage drop reduces the plate voltage applied to the audio tube with a consequent reduction of its amplification.

When a signal is tuned in, an avc voltage is developed which biases the suppressor tube to cutoff. No plate current for this tube, therefore, flows through the resistor and the receiver and avc circuit function in a normal manner.

**Noise Currents.** Free electrons in a good conductor wander from one atom to another and, owing to their thermal energy, produce random and haphazard currents within the interior of the conductor, even though no voltage difference is applied to the conductor. This random movement of electrons results in no useful current, because as many electrons in any small region of the conductor move in one direction as in the opposite direction during an increment of time

At any instant, however, a net movement of electrons may be in one direction, and at the next instant a net movement may exist in the opposite direction, giving rise to an alternating current within the conductor. Since this type of movement is purely haphazard in its alternations, the resulting current is known as **noise current,** or simply **noise,** because these effects are of utmost importance in amplifying very weak signals such as those generated by microphones and similar sources. Noise currents impose a lower limit on the strength of a signal that can be amplified, because, if the desired signal is not sufficiently greater in amplitude than the noise signal, the output of the amplifier will be unintelligible

## *THE FREQUENCY-MODULATION RECEIVER*

In order to understand frequency modulation receivers, it is necessary to have a clear conception of the nature of a frequency-modulated carrier wave. In the conventional form of modulation (amplitude modulation), the carrier envelope is varied in *amplitude* in accordance with the wave form of the impressed modulating a f signal. In frequency modulation, the carrier is varied in *frequency* at a *rate* corresponding with the frequency of the impressed modulating audio frequency. The *extent* of this variation in carrier frequency depends upon the amount of modulation it is desired to produce.

Thus, if it is desired to transmit a 500-c a-f sound wave by frequency modulation of a 1,000-kc carrier wave, this may be accomplished as

follows: The 1,000-kc carrier is modulated by varying its frequency between 1,000,010 and 999,990 c 500 *times per second.* If it is desired to transmit a 700-c audio tone, this may be accomplished by varying the 1,000-kc carrier between 1,000,010 and 999,990 c 700 *times per second.* It is seen, therefore, that the frequency of the audio signal obtained from a frequency-modulation receiver tuned to the above signal depends upon the *rate* at which the carrier varies in frequency between the limits

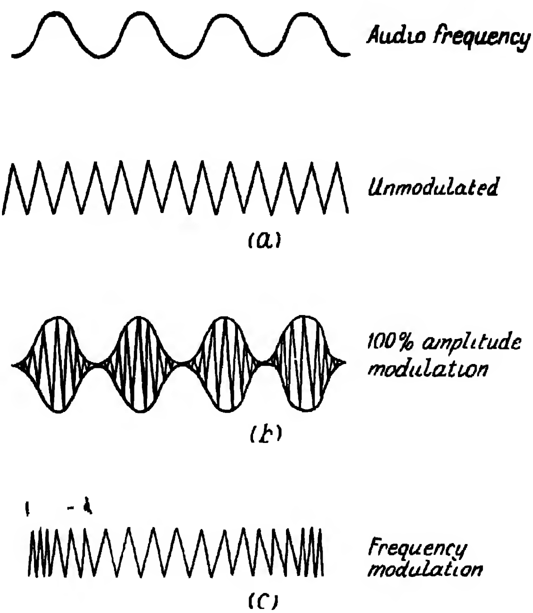



FIG. 176. Comparison of frequency modulation and amplitude-modulation wave forms.

mentioned above. The *intensity* of the received audio signal is a function of the band width over which this frequency variation occurs. Thus, in the above illustration, the band width over which the variation in frequency occurs is only 20 c. If it were desired to double the intensity of the transmitted signal, it would be necessary to vary the frequency *at the same rate* over twice the band width, or 40 c. Thus, to double the intensity when modulating the 1,000 kc carrier with the 500-c signal, it would be necessary to vary the carrier frequency between 1,000,020 and 999,980 c at the rate of 500 c per second.

Figure 176 illustrates the graphical wave form of a frequency modulated signal as compared with an amplitude-modulated signal when both are modulated from the same source. It should be understood that the band width figures mentioned in the above example are not representative but were used merely to illustrate the point. Actually, good modulation

when using the system of frequency modulation requires the use of a 160-kc band where noise-free reception is desired. As a matter of fact, the extremely wide band required is responsible for the confinement of frequency-modulation transmission to the u-h-f portion of the radio spectrum. Such wide bands could not be tolerated on the already over-crowded present-day broadcast bands.

**Principle of Operation.** The essential difference between frequency-modulation receivers and amplitude-modulation receivers, so far as the principle of operation is concerned, lies in the method of detection. Detection, as was seen in an earlier part of this chapter, consists fundamentally of separating the a-f components of a received signal from the carrier frequency. In amplitude modulation receivers, detection can be thought as converting variations in *amplitude* of the carrier to variations in amplitude of an a-f signal. In frequency modulation receivers, detection can be thought of as converting variations in the *frequency* of the carrier to variations in amplitude of an a-f signal.

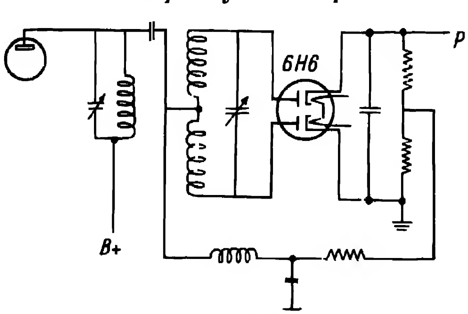


FIG. 177. Frequency discriminator circuit in frequency-modulation receiver

In amplitude modulation detector systems, the *frequency* of the audio output depends upon the *frequency* of the variations in carrier amplitude. The *amplitude* of the audio output in such systems, depends upon the *amplitude* of the carrier variations. In frequency modulation detector systems, the *frequency* of the audio output depends upon the *rate* (variations per second) at which the carrier varies in frequency between the limits of the system. The *amplitude* of the audio output depends upon what these limits are, in other words, upon the band width. The wider the band, the greater the amplitude of the audio output from the detector.

The detector in a frequency modulation receiver is usually called the **discriminator,** or **discriminator detector.** This distinguishes it from the conventional types of detection and is representative of the actual function of the circuit, that is, frequency discrimination.

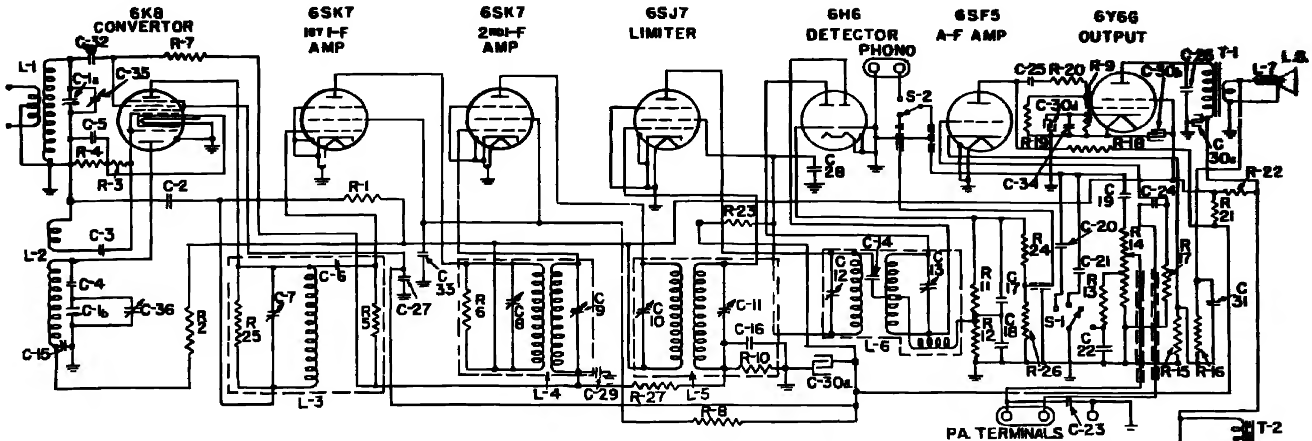
A typical discriminator circuit is shown in Fig. 177. The special i-f transformer has its secondary center tapped with each half of the winding operating one diode of a double-diode tube. The primary of this transformer is tuned to the exact intermediate frequency of the receiver. When a frequency-modulated signal is received, the variation in i-f frequency, of course, occurs in exact accordance with the variations in carrier frequency. One half of the i-f transformer secondary winding

is tuned to a frequency somewhat higher than the intermediate frequency. The other half of the secondary winding is tuned to a frequency lower than the intermediate frequency. The resistors $R_1$ and $R_2$ across the cathodes represent the load on the respective diode sections of the tube.

When the i-f signal does not vary in frequency (no modulation), the signal voltages developed across the diode load resistors will be equal and opposite in polarity. Since the voltage fed into the audio amplifier is taken between point $P$ (Fig. 177) and ground, zero voltage will be developed across the audio input circuit under these conditions. When the i f signal is frequency modulated, each diode in turn will develop a greater voltage across its load resistor than the other. At any instant, the voltage developed across the audio-amplifier input circuit will be the algebraic sum of the oppositely polarized voltages developed across the individual load resistors.

This action can be more clearly followed if the voltages developed by a typical signal are analyzed for a single cycle of audio modulation. Assume that with no modulation, the carrier develops 50 v across each diode load resistance. The total output audio voltage will therefore be zero, since the individual diode load voltages are opposite in polarity. When the signal is modulated the carrier shifts alternately to higher and lower frequencies. When the carrier has *increased* in frequency, the voltage developed across the load resistor of the diode whose circuit is tuned to a *higher* frequency *increases*, possibly to 65 v. The voltage across the load resistor of the diode whose circuit is tuned to a *lower* frequency *decreases* by a similar amount, thus falling to 35 v. The total voltage developed across the following stage of audio amplification is the algebraic sum of these two voltages, or 30 v It follows that the amplitude of the a-f voltage developed in the detector output circuit is a function of the extent of the variation in carrier frequency. This is the reason for the wide band width used in frequency modulation systems. The wider the frequency band used, the greater the amplitude of the audio output voltage developed by the discriminator circuit. The above voltages have been arbitrarily chosen for the purpose of illustration and are not intended to be representative.

On the opposite frequency shift of the mythical signal under discussion, the carrier *decreases* in frequency. The *higher* frequency diode circuit now develops only 35 v across its load resistor, and the *lower*-frequency diode circuit develops 65 v across its load resistor. The total, or net, output voltage developed is again 30 v, but this time is of opposite polarity. The variation in frequency of the carrier has therefore caused an output voltage to be developed that changes in polarity, that is, an *alternating* voltage. Since the polarity of the output voltage depends upon the *direction* of carrier-frequency change (whether *increasing* or *decreasing*), it follows that the *frequency* of the audio signal thus developed



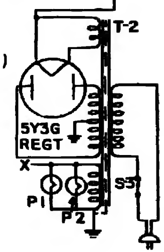



FIG. 178. Typical frequency modulation receiver circuit. Parts are identified in the list below. (Courtesy of General Electric Company.)

| Symbol | Description |
|---|---|
| Capacitors | |
| C-1 | 20 μμf trimmer |
| C-2 | .02 μf 400 v paper |
| C-3 | 47 μμf mica |
| C-4 | 1,500 μμf mica |
| C-5 | .01 μf 200 v paper |
| C-[illegible] | 300 μμf trimmer |
| C-14 | 47 μμf mica |
| C-15 | 470 μμf mica |
| C-16 | 22 μμf mica |
| C-17 | 100 μμf mica |
| C-18 | 100 μμf mica |
| C-19 | .005 μf 400 v paper |
| C-20 | .002 μf 600 v paper |
| C-21 | 470 μμf mica |
| C-22 | .002 μf 600 v paper |
| C-23 | 220 μμf mica |
| C-24 | .007 μf 400 v paper |
| C-25 | .05 μf 400 v paper |
| C-26 | .005 μf 600 v paper |
| C-27 | .005 μf 400 v paper |
| C-28 | .05 μf 400 v paper |
| C-29 | .05 μf 200 v paper |
| C-30a | 20 μf 250 v dry electrolytic |
| C-30b | 20 μf 250 v dry electrolytic |

| Symbol | Description |
|---|---|
| Capacitors (cont.) | |
| C-30c | 40 μf 2[illegible] v dry electrolytic |
| C-30d | 20 μf [illegible] v dry electrolytic |
| C-31 | 0.1 μf 200 v paper |
| C-32 | 47 μμf mica |
| C-33 | 0.1 μf 400 v paper |
| C-34 | [illegible] μf [illegible] v paper |
| C-35 | 2 to 12 μμf [illegible] trimmer |
| C-36 | 7 to 2[illegible] μμf [illegible] trimmer |
| Resistors | |
| R-1 | 2,200 ohms ½ w carbon |
| R-2 | 1,500 ohms ½ w carbon |
| R-3 | 47,000 ohms ½ w carbon |
| R-4 | 350 ohms ½ w carbon |
| R-5 | 470,000 ohms ½ w carbon |
| R-6 | 47,000 ohms ½ w carbon |
| R-7 | 470,000 ohms ½ w carbon |
| R-8 | 4,700 ohms ½ w carbon |
| R-9 | 220 ohms 2 w carbon |
| R-10 | 330,000 ohms ½ w carbon |
| R-11 | 100,000 ohms ½ w carbon |
| R-12 | 100,000 ohms ½ w carbon |
| R-13 | 120,000 ohms ½ w carbon |
| R-14 | 2 meg [illegible] volume control |
| R-15 | [illegible] ohms ½ w [illegible] |

| Symbol | Description |
|---|---|
| Resistors (cont.) | |
| R-16 | 2,200 ohms ½ w carbon |
| R-17 | [illegible] ohms ½ w carbon |
| R-18 | 220,000 ohms ½ w carbon |
| R-19 | 470,000 ohms ½ w carbon |
| R-20 | 1,500 ohms ½ w carbon |
| R-21 | 2,200 ohms 1 w carbon |
| R-22 | 1,600 ohms 4 w wire wound |
| R-23 | 2,200 ohms ½ w carbon |
| R-24 | 100,000 ohms ½ w carbon |
| R-25 | [illegible],000 ohms ½ w carbon |
| R-26 | 250,000 ohms ½ w carbon |
| R-27 | 2.2 megohms ½ w carbon |
| Switches | |
| S-1 | tone switch |
| S-2 | phono switch |
| S-3 | power switch (on S-2) |
| Transformers | |
| L-1 | antenna |
| L-2 | oscillator |
| L-3 | first I F |
| L-4 | second I F |
| L-5 | I F limiter |
| L-6 | discriminator |

depends upon the *rate* at which the carrier frequency varies between its band-width limits.

So far as fundamental principles are concerned, the essential difference between frequency modulation and conventional receivers lies in the detector, or discriminator, circuit. There are several other differences, however, which, although they represent no departure from the principle of operation of conventional receivers, *do* constitute a difference in design.

The importance of large band width in frequency modulation receivers has already been discussed. In order to permit the impression of so wide a band of frequencies upon the discriminator circuit, it is necessary that the preceding stages permit the passage of all frequencies within this band with uniform amplification. The design of frequency-modulation receivers has been standardized to provide a band width of 200 kc, and accordingly, all frequency-modulation receivers are designed so that r-f and i-f stages provide essentially linear response over a band 100 kc either side of the carrier. Little difficulty is experienced in obtaining this response in the r-f stages at the frequencies at which frequency-modulation receivers operate. A flat top on the frequency response curves of i-f amplifier stages is obtained by introducing resistance in the primary of the i-f transformers. Values of about 15,000 ohms are common. In all other respects, the i-f transformers are of conventional double tuned design and do not differ fundamentally from standard units.

It is essential for the proper functioning of the discriminator circuit that all frequencies within the band limits are of equal amplitude at the discriminator input circuit. Any variation of amplitude of the frequency-modulated i-f signal at different frequencies will result in variations of the voltages developed across the diode load resistors. Such variations, since they form no part of the original signal, appear as distortion, or interference, in the a-f amplifier. Thus, if the signal amplitude at the discriminator input circuit is greater at the high frequencies than at the low, owing to nonuniform amplification, higher voltages are developed across the h-f diode load resistor than across the l-f diode load resistor. As a result, the wave form of the a-f voltage developed in the discriminator is distorted.

In order to prevent such amplitude distortion, the stage immediately preceding the discriminator stage is operated at zero initial bias and comparatively low plate voltage. Plate-current cutoff therefore occurs at relatively low values of negative signal input. Positive signal impulses cause grid current to flow. The voltage drop across a resistor placed in the ground return lead prevents the grid voltage from following the positive peaks. All variations in amplitude that exceed a certain limit cause this tube to cut off either on the positive or negative swing of the input cycle. This circuit, called the **limiter**, applies a constant amplitude signal to the discriminator.

One of the outstanding advantages of the frequency-modulation system of transmission is the comparative freedom from noise. Practically all man-made and natural (static) electrical noises appear in a receiver as a form of amplitude modulation. Amplitude peaks caused by such noise pulses are effectively cut down to the signal level by the limiter stage. Therefore, they do not affect the discriminator stage and, hence, do not appear in the audio output. To a small extent, some forms of such noise affect the frequency modulation by appearing as pulses *between* cycles of the signal. They thus momentarily create an effective increase in band width. As previously discussed, the amplitude of the audio voltage developed by the discriminator circuit is a function of the band width, and noises of this nature appear as momentary fluctuations in amplitude in the audio output. This type of noise is discriminated against by using wide band carrier shift. As a result, the combined effect of the limiter circuit and large band width is to produce a signal-to-noise ratio unheard of in amplitude-modulation receivers.

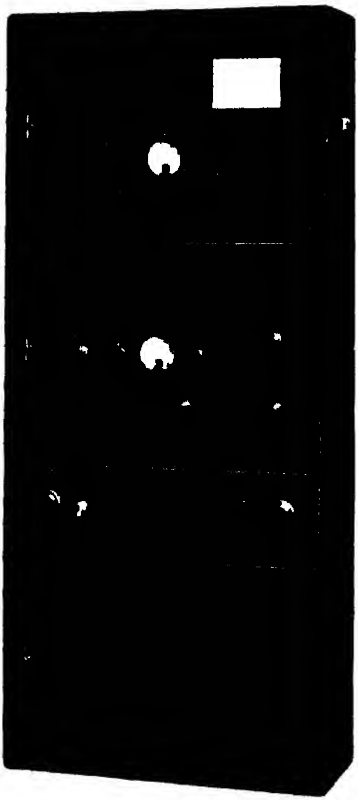


FIG. 180(a). Typical marine auto alarm front view (*Courtesy of Mackay Radio & Telegraph Co., Inc.*)

The conception of frequency modulation is not entirely new, but until about 1935, its real advantages were unknown. At this time, Major E. H. Armstrong demonstrated the advantages by using wider frequency swing than had been used previously and incorporating an effective amplitude limiter in his receiver. This combination greatly improved the signal-to-noise ratio.

The advisability of using a system of radio transmission that requires a band width of 200 kc has been questioned, since a shortage of space is already apparent in the broadcast spectrum. However, the development of frequency modulation systems will not upset the present broadcast system, because all operations are in the u-h-f range. As a matter of fact,

space may actually be conserved. Because of the limited range in the u.h.f. spectrum, stations only a few miles apart can operate on the same frequency with very little interference area. In addition, owing to the effective discrimination against both natural and man-made static,

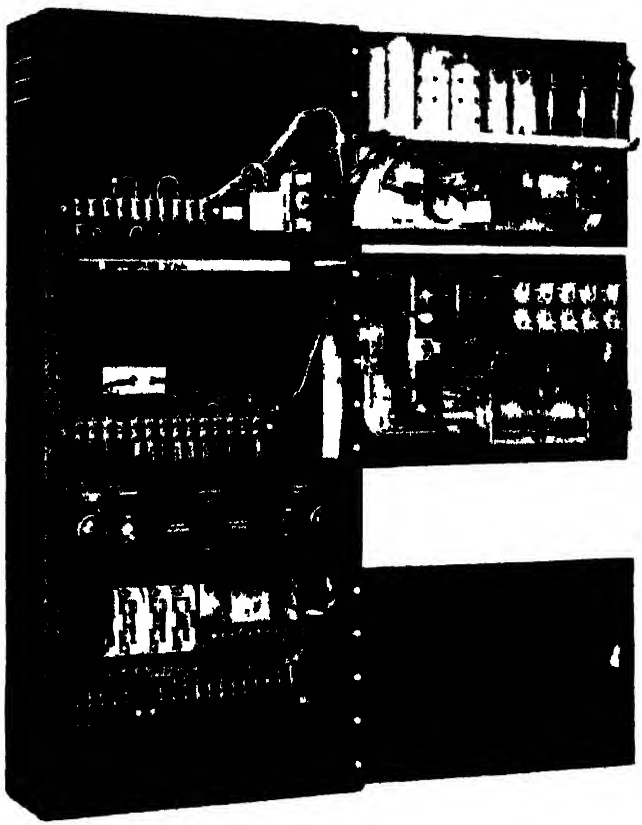


Fig. 180(b). Typical marine auto alarm, front with panels swung open. (*Courtesy of Mackay Radio & Telegraph Co.*)

reliable reception is obtained up to the critical distance beyond which the field intensity falls off.

### *THE AUTO ALARM RECEIVER*

Auto alarm receivers are safety devices that automatically ring an alarm bell upon the reception of a special signal called the international

**auto-alarm signal.** In general, an auto-alarm system consists of a conventional receiver, either tuned r-f or superheterodyne, in connection with a selector that ensures that the alarm relays are set off only by the proper combination of incoming signals. The receiver is fixed-tuned on the international distress frequency of 500 kc.

The international auto-alarm signal adopted by international law consists of a series of 12 dashes of 4 sec duration with 1-sec intervals between the dashes. International law requires that a ship in distress transmit this auto-alarm signal *prior* to the distress signal SOS. Ships having only one operator are required to have auto-alarm receiving systems in operation at all times when the operator is not on watch. The reception of the auto alarm signal described above operates a stepping relay, which is really a form of rotary switch equipped with a ratchet. Any received signal of the proper duration will cause the stepping relay to be advanced one position. Sufficient leeway is provided in the timing of the apparatus so that a signal from 3.5 to 4.5 sec in duration will operate the ratching relay.

A minimum of four received signal pulses of the proper duration and spacing will operate the alarm system. When the stepping relay advances four positions, it closes a pair of contacts that actuate the alarm system. Another relay is so wired into the circuit that any excessive delay between signal pulses (longer than 1 sec) will actuate it to return the stepping relay to zero position. Thus, only dashes of the proper length can actuate the stepping relay, and only those properly spaced can continue to operate it.

When the stepping relay is closed by four properly spaced dashes of the correct length, it closes an alarm bell circuit. At the same time, a circuit is opened, which prevents further signals from returning the stepping relay to zero position.

The alarm bells are customarily located in the operator's quarters, on the bridge, and in the radio room itself. Thus, if for some reason, the operator is not recalled to his post by the alarm signal, he will be notified by the officer on watch on the bridge. Most installations also include a red warning light which operates whenever the stepping relay is advanced one position. The operator is thus advised whenever heavy static is impairing the operation of the instrument and can accordingly adjust the sensitivity of the receiver

## *QUESTIONS AND PROBLEMS**

**1.** What are the primary characteristics of a choke input filter?

**2.** Is a grid-leak type of detector more or less sensitive than a power detector (plate rectification)? Why?

* These questions and problems are taken from the "F C C. Study Guide for Commercial Radio Operator Examinations."

**3.** What are the advantages to be obtained from adding a tuned r-f stage ahead of the first detector (converter) stage of a superheterodyne receiver?

**4.** How is automatic volume control accomplished in a radio receiver?

**5.** If a superheterodyne receiver is tuned to a desired signal at 1,000 kc and its conversion oscillator is operating at 1,300 kc, what would be the frequency of an incoming signal which would possibly cause image reception?

**6.** Why does a screen-grid tube normally require no neutralization when used as an r-f amplifier?

**7.** In a class A a-f amplifier, what is the main advantage obtained through the use of two triodes in push-pull as compared to parallel operation?

**8.** Explain how power detection is accomplished.

**9.** What is the function of the grid leak in a grid-leak type of detector?

**10.** What are the advantages of push-pull amplification as compared to single-ended amplification?

## Chapter XIII

# TRANSMITTING-CIRCUIT PRINCIPLES

Modern transmitting systems may be divided into three parts: the power supply, the oscillator system, and the amplifier. In broadcast or radiotelephone transmitters, a fourth part comprising a speech-amplifier and speech-modulator system may be added. Each of these divisions of a radio transmitter will be discussed separately.

### *THE POWER SUPPLY*

Transmitter power supplies may, in general, be placed in two major classifications, namely, those utilizing alternating current as a basic source of power and those utilizing direct current as the power source. Usually, where only direct current is available for power supply purposes, some form of motor generator having an output of the proper voltage and current rating is used to provide the direct current necessary for the operation of the transmitter vacuum tubes. Most shipboard installations fall under this classification, although some such installations utilize motor generators with an output of 110 or 220 v alternating current, either single phase or polyphase. This alternating current is then stepped up in voltage by a transformer, rectified, and filtered.

Many broadcast and coastal telegraph shore stations utilize d c generators driven from an a c source to supply power to transmitter vacuum tubes. These installations usually include a number of generators to supply the necessary power for plate circuits, bias circuits, and filament or heater operation.

Motor generators utilized to fulfill the above functions have been treated in Chap. VI and will therefore not be discussed here.

The most widely used source of power for radio transmitters is the alternating current supplied by power companies. Power supplies operating from such sources are classified according to the nature of the alternating current concerned. There are usually two such classifications — the *single-phase system* and the *three-phase system* and it is customary to refer to any a-c system having more than a single phase as a **polyphase system.** Although two-phase systems are utilized for some forms of power transmission, they are not in general use and are not used in transmitter power-supply systems.

**Single-phase Power Supplies.** The single-phase power supplies used

in transmitters are identical in principle with the power supplies discussed for receivers in the beginning of Chap. XII. Of course, power supplies for most transmitters must deliver a much higher voltage than receiver power packs. Nevertheless, aside from the voltage and current demands, the two types of power supplies are identical. All the design data and theory that were discussed in Chap. XII can be applied with equal accuracy to transmitter systems.

In some respects, fundamental design practice is altered in transmitter single-phase power-supply systems. For example, the exceedingly large transformers required to supply high voltage to the rectifier plates can be more efficiently manufactured as separate components. Separate transformers are utilized to supply filament or heater current. Similarly, a number of entirely separate power supplies are used in high-power transmitters to provide power for individual stages of the transmitter. This practice is usually economical as well as efficient. It is easier to design an efficient power pack to supply a fixed voltage with relatively constant load than to design a power pack to supply a number of different voltages with resultant loss of efficiency. Furthermore, power-output stages require less filtering than do the earlier stages in a transmitter. Filter components (capacitors and chokes) for the tremendously high voltages used in some of these stages are expensive. Although the amount of filtering required in earlier stages of a transmitter is greater, the filter components for these lower-voltage power supplies are comparatively inexpensive.

The use of separate power supplies for separate stages of a transmitter offers other advantages. Radio-frequency feedback from the output stage to an earlier stage through the common impedance of the power supply is completely eliminated and obviates the necessity for r-f filtering and decoupling networks, which would have to be used with a common supply. $L$ and $C$ components for such networks are rather expensive items in high-power stages.

Separate power supplies to supply bias power are practically always utilized in transmitters of appreciable power. Aside from the practical advantages, there are several important advantages of this practice, which will be discussed in the circuit section of this chapter.

Single-phase rectifier and filter systems in transmitter power packs do not differ radically in practical design from receiver power packs. Of course, rectifier tubes capable of handling greater currents at higher voltages are required. Mercury-vapor tubes are more often used in transmitters because of the small voltage drop in the tube.

**Polyphase Power Supplies.** Polyphase alternating currents of three phases are very commonly used as the primary power source for radio-transmitter power supplies. A three-phase source is one comprised of three distinct emfs. The three phase currents used for transmitter

circuits and supplied by municipal power companies are *balanced*. Such a three-phase source is said to be balanced when the three emfs have the same effective values and when they are displaced in phase from each

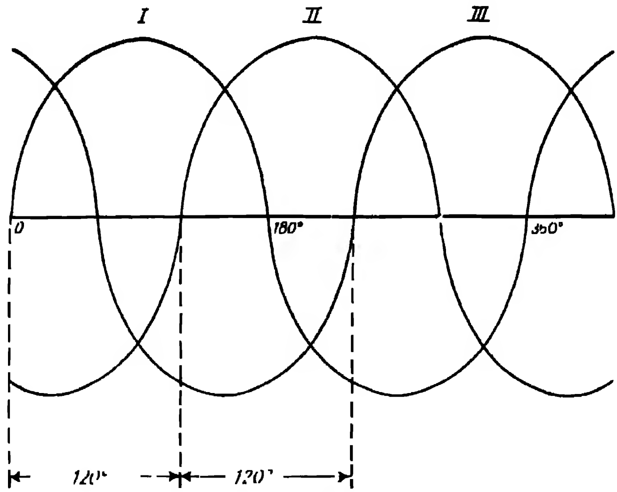



FIG. 181. Three-phase alternating current.

other by 120°. The condition of perfect balance is assumed in the following discussion.

A graph of a three-phase alternating current is shown in Fig. 181. It

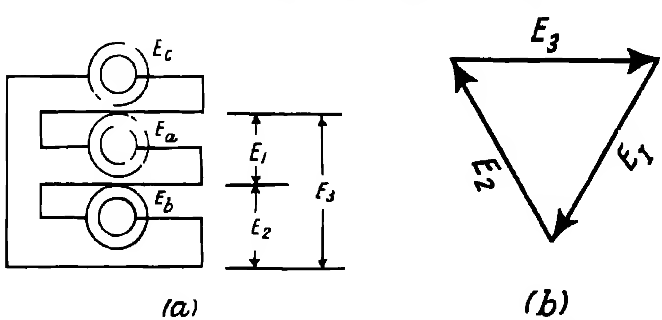



FIG. 182. Delta-connected three-phase system. (a) Three-phase a-c source connected in delta. (b) Vector diagram of emfs in delta-connected system.

is apparent from this graph that at any instant the net voltage (the vector sum of the individual voltages) is zero. Hence, the terminals of the three emfs may be connected in series, and there will be no residual voltage in the combination. Such a connection is known as the "delta," or "mesh," connection. A diagram of a delta-connected system with a vector diagram representing the emfs is shown in Fig. 182. This

connection receives its name from the resemblance of the vector diagram to the Greek letter Δ (delta).

A diagram of a Y- (wye) or star-connected three-phase system is shown in Fig. 183 with its accompanying vector diagram. This connection receives its name from the resemblance of its vector diagram to the letter Y or to a three-cornered star. Although the diagrams of Figs. 182 and 183 show independent sources for each phase emf, these sources are simply so shown for ease of illustration. Actually, modern power plants generate three-phase alternating currents by means of revolving field

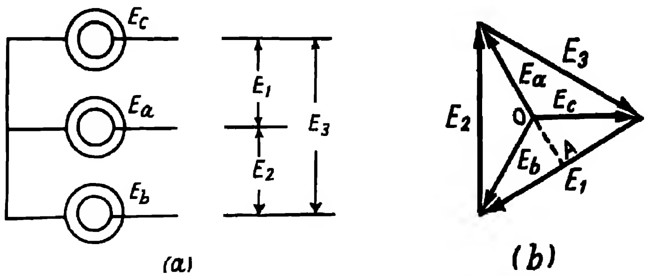



FIG. 183 Wye-connected three-phase system (a) Three-phase a-c source connected in wye (b) Vector diagram of emfs in wye-connected system

alternators. Such alternators have three distinct armature windings which are either delta- or star-connected.

In the star-connected system of Fig. 183, it is seen that

$$E_1 = E_a - E_c, \tag{1}$$

$$E_2 = E_b - E_a, \tag{2}$$

$$E_3 = E_c - E_b. \tag{3}$$

Adding Eqs. (1), (2), and (3),

$$E_1 + E_2 + E_3 = E_a + E_b + E_c - E_c - E_a - E_b, \tag{4}$$

and

$$E_T = 0, \tag{5}$$

where $E_T$ = net line voltage.

By dropping a perpendicular ($OA$, Fig. 183) from the $O$ point to one of the line-voltage vectors, the line-voltage vector is bisected. Then, by trigonometry,

$$\frac{E_L}{2} = E_p \sin 60°, \tag{6}$$

where $E_L$ = line voltage = $E_1$, $E_2$, or $E_3$,
$E_p$ = phase voltage = $E_a$, $E_b$, or $E_c$.

From Eq. (6),

$$E_L = 2E_p \sin 60°, \tag{7}$$

or

$$E_L = 2E_p\,(0.866), \tag{8}$$

and

$$E_L = 1.73E_p. \tag{9}$$

In the delta-connected system of Fig. 182, it is apparent that

$$E_1 = E_a, \tag{10}$$

$$E_2 = E_b. \tag{11}$$

$$E_3 = E_c. \tag{12}$$

However, it can be shown that the vector relations between the phase and line *currents* in a delta connected system are the same as the vector relations between the phase and line *voltages* in a star-connected system. It follows, therefore, that in a delta-connected system

$$I_L = 1.73I_p, \tag{13}$$

where $I_L$ = line current;
$I_p$ = phase current.

From the foregoing, it will be seen that the power in a star-connected three-phase system is

$$P = 1.73E \cdot I \cdot \cos \theta, \tag{14}$$

where $P$ = true power of the circuit;
$E$ = line voltage across each phase;
$I$ = line current in each phase line;
$\cos \theta$ = power factor of the circuit.

Similarly, the power in a delta-connected three-phase system is

$$P = E \cdot 1.73I \cdot \cos \theta, \tag{15}$$

where the symbols have the same meaning as in Eq. (14).

For general use, the equation

$$P = E \cdot I \cdot 1.73 \cdot \cos \theta \tag{16}$$

can be applied either to delta or star-connected systems since numerically Eqs. (14) and (15) are equivalent.

As in single-phase circuits, the basic unit of the polyphase-transmitter power supply is the plate, or high-voltage, transformer. This transformer, in addition to performing the very necessary function of voltage step-up, may also be considered as a coupling device used to connect the polyphase voltage source to a polyphase-rectifier load.

There are three methods of coupling three-phase voltage sources to polyphase loads by means of transformers. The first of these is called

the **delta-delta connection** and consists of connecting the primary and secondary windings, respectively, in delta, as shown in Fig. 184(a). This system has the main advantage that three similar single-phase transformers can be so connected instead of utilizing a single multiwinding unit. Furthermore, if one unit is disabled, the bank can be connected open-delta. This connection, often also called **V-V connection,** is shown in Fig. 184(b). Although the capacity of the system when so connected

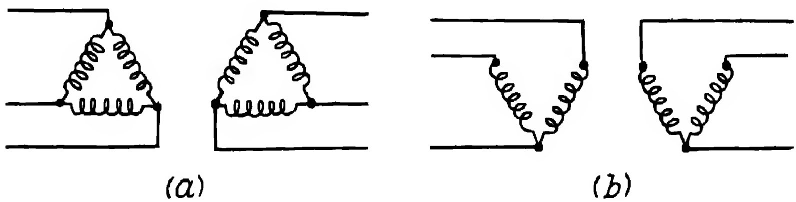


FIG. 184. Three-phase transformer connections. (a) Delta-delta transformer connection. (b) Open-delta (V-V) transformer connection.

falls to approximately 58 per cent of the three-unit capacity, it is useful in an emergency.

For low-voltage, high-current service, the delta-delta connected system is an economical design and is also free from third harmonic troubles. For radio application, it has the serious disadvantage that the neutral cannot be derived, and in high voltages the design is expensive.

The second method of connecting polyphase transformers is called the

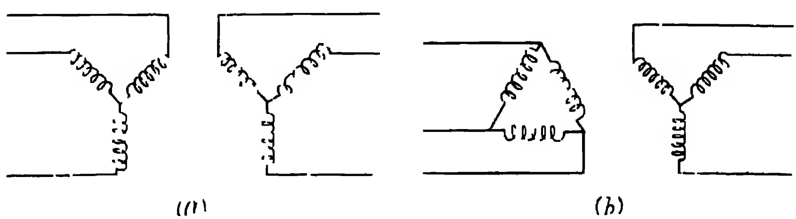


FIG. 185. Three-phase transformer connections. (a) Star-star (Y-Y) transformer connection. (b) Delta-star transformer connection.

**Y-Y method** and is shown in Fig. 185(a). Among the desirable features of this type of connection is the fact that the neutral can be brought out for grounding. For high-voltage, low-current applications, the design is more economical than the delta. In addition, a short circuit in or on one unit does not cause a power short circuit, although it raises the voltage on the other units to 1.73 times normal voltage. The main disadvantage of this unit is the very large third-harmonic content in the voltage of each phase if the neutral of the primary is not properly grounded.

The third method of transformer connection is called the **delta-Y, or delta-star,** connection. This is generally considered to be the most

satisfactory connection and is the one used in practically all radio-transmitter power supplies. The neutral can be brought out for either grounding or loading and is very stable, and the connection is practically free from third-harmonic voltages. Differences of magnetizing current, voltage ratio, and impedance in the different units are adjusted by a small magnetizing current circulating in the delta. This type of connection is shown in Fig. 185(b). The delta-star connection, however, cannot operate temporarily with two units when one is disabled. A short circuit in one unit is extended to all three units. Also, if the delta on the primary side is accidentally opened, the unexcited leg on the Y side may resonate with the line capacitance and cause damage.

**Polyphase Rectifier Systems.** Polyphase rectifying circuits may be

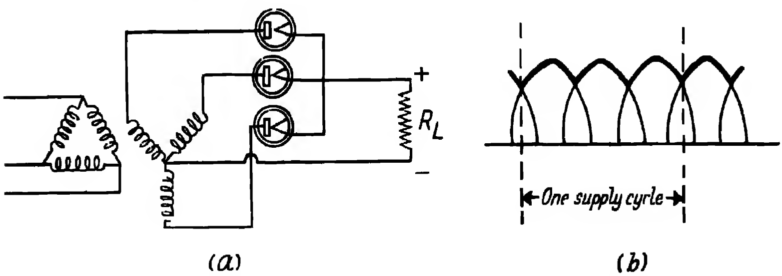



FIG. 186. (a) Three-phase half-wave rectifier circuit. (b) Output wave form of (a).

compared in type to single-phase circuits. As in single-phase circuits, polyphase rectifiers are of two types, namely, full-wave rectifiers and half-wave rectifiers. Because of the much smaller ripple component output of any style of polyphase rectifier system, the amount of filtering required is much less than with single-phase rectifiers. Even a half-wave polyphase rectifier has a much smaller output ripple component than a full-wave single-phase rectifier. This is an important feature of polyphase circuits and is responsible for the use of polyphase rectifier systems in practically all transmitters with a power rating of over 1 kw.

Figure 186(a) illustrates a typical three-phase half-wave rectifier circuit. This circuit consists essentially of three half-wave rectifier tubes, each connected to one leg of the Y-type secondary winding. The individual phase circuits are much the same as a corresponding half-wave single-phase rectifier, the common neutral forming the negative return lead. In this arrangement, each rectifier tube carries current one-third of the time. The output wave shown in Fig. 186(b), therefore, pulsates at three times the frequency of the a-c supply. For conventional 60-c supply sources the ripple component will be 180 c. In half-wave rectifier systems of this type, the ripple frequency component in the output (filter circuit input) will be on the order of 50 per cent.

A three-phase full-wave rectifier circuit is shown in Fig. 187(a). With this system, full-wave rectification is obtained through each leg of the secondary winding. This circuit is capable of delivering much higher voltage output than a half-wave system utilizing the same transformer. Although four separate filament transformers and six rectifier tubes are required, compared with one filament transformer and three tubes of the half-wave system previously described, the full wave is generally preferred by engineers because of its greater efficiency.

A graph of the output wave of a three-phase full-wave rectifier is shown in Fig. 187(b). The pulsations of the output current occur at a rate six times the frequency of the a-c. source. For 60-c circuits, therefore, the output ripple frequency is 360 c. The ripple-frequency component in the

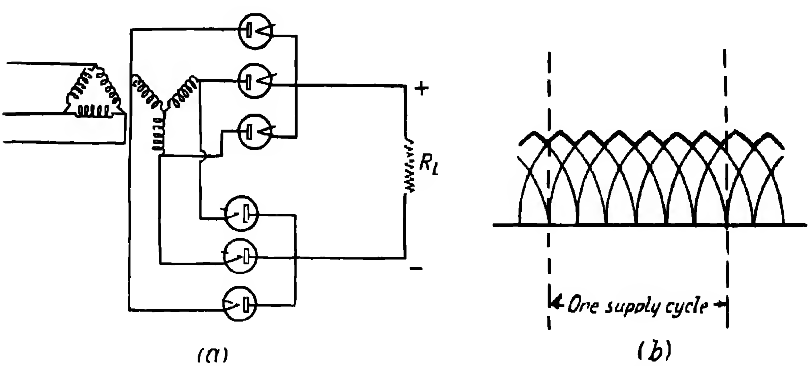



FIG. 187 (a) Three phase full wave rectifier circuit (b) Output wave form for (a).

output of three phase full-wave rectifier systems is of the order of 14 per cent and is very easily filtered.

Another system very commonly used in transmitter power supplies is the three-phase half-wave double Y rectifier circuit shown in Fig. 188(a). This consists essentially of two three-phase half-wave rectifiers with their outputs connected in parallel. The phase relation between the two systems is such that when the output of one unit is at minimum, the output of the other is at maximum. Consequently, the ripple frequency and the percentage of ripple-frequency component in the output are substantially the same as for the full-wave system described above.

The interphase reactor, or balance coil, serves the purpose of equally dividing the current between the two units. Each half-wave three-phase unit is therefore enabled to operate independently. If the reactor were not in the circuit, each tube would carry the load current only one sixth of the time. The proper current distribution maintained by the balance coil permits each tube to operate one third of the time. Therefore, at any instant, there are two tubes delivering current to the load. As a

result, this circuit is capable of delivering twice the current of the half-wave circuit of Fig. 186 with a ripple frequency of the full-wave circuit of Fig. 187. Although this circuit will not deliver as great a voltage output as the full-wave three-phase system, it requires only one filament transformer and is used where heavy current demands must be met.

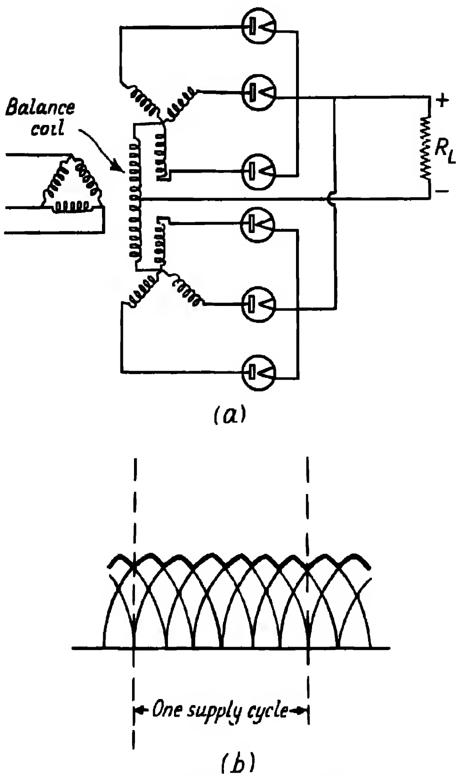



FIG. 188 (a) Three-phase half-wave double-Y rectifier circuit (b) Output wave form of (a)

**Polyphase Filter Systems.** Filter systems for polyphase rectifier circuits must, in general, meet the same requirements as those of single-phase circuits. Since the ripple-frequency component is so much smaller and the ripple frequency so much higher, polyphase filters are relatively small, that is, the components have comparatively small values of inductance and capacitance. All the formulas developed in Chap. XII may be applied in designing filters for polyphase rectifiers. It should be remembered, however, that in polyphase systems, one does not deal with input ripple components of 100 per cent as in single-phase systems. The percentage of input ripple component to the first filter section will depend upon the type of rectifier system used, as described above.

## THE TRANSMITTER OSCILLATOR

Any oscillator in itself may be considered a complete transmitter. If it were coupled to an antenna system, it would radiate the signal that it generates in as efficient a manner as a complete multistage transmitter. A number of factors, however, limit the use of simple oscillators as transmitters. In the first place, it is difficult to construct an oscillator of appreciable power output and maintain a reasonable circuit stability. Such an oscillator, too, would have its circuit constants disturbed by the

variable reflected load presented by the antenna system. In addition, in order to deliver the power output required for modern radio transmission, single oscillators would require the use of vacuum tubes so large that their manufacture would be unfeasible.

Modern design practice is to use small vacuum tubes as oscillators. Even high-powered broadcast transmitters utilize small receiving tubes of the type employed in receiver power-output stages for the oscillator stage. A small oscillator of this type can more easily be efficiently designed for stable operation.

Practically all modern transmitters utilize crystal oscillators because of the frequency-stabilization feature. As a matter of fact, broadcast and fixed-service telegraph and telephone transmitters are required by law to use crystal control. Many ship radiotelegraph transmitters utilize self-excited oscillators, but present-day tendencies are gradually to eliminate these. Eventually, all transmitters in use—even ship transmitters utilizing a number of different frequencies—will be crystal-controlled.

An oscillator converts d-c energy supplied by the plate battery (or power pack) into a-c energy, which is fed to the plate tank circuit. The frequency at which oscillations occur is determined by the $L$ and $C$ constants of the tank circuit, but the shunting effect of the tube capacitances and reflected impedances of the grid and load coupling circuits also considerably affect the tank circuit parameters. In the case of a crystal oscillator, the frequency at which the tube will oscillate is limited by the natural period of the crystal. Tuning of the oscillator circuit, however, is controlled by the tank circuit. When the tank-circuit frequency approaches the natural frequency of the crystal, oscillations commence.

Oscillator conversion efficiencies higher than the theoretical maximum of 50 per cent for class A operation are obtained by biasing the grid sufficiently negative so that plate current flows during only a portion of the cycle. The plate voltage with respect to the cathode during the period in which plate current flows is at a minimum, so the power loss in the tube is small. The oscillator thus operates as a type of class C amplifier, and efficiencies of 80 and 90 per cent may be obtained. The general theory of vacuum-tube oscillators has been discussed in Chap. XI and will not, therefore, be further enlarged upon here.

In marine installations employing self-excited oscillators, provision is made for operation on a number of frequencies. This is usually accomplished by taps on the main oscillator tank inductance. Vernier adjustment is obtained by continuously variable additional inductances. The FCC requires that the frequency at each frequency for which the ship station is licensed be checked at periodic intervals and recalibrations made if necessary. In marine crystal-controlled transmitters, flexible operation is procured by means of separate crystals for each licensed

frequency. Oscillator tuning is a matter of simply switching the proper crystal into the circuit. The same switch arm simultaneously cuts in the proper amount of inductance or capacitance in the oscillator tank circuit.

Three tuning adjustments are necessary before operating a marine radiotelegraph transmitter: The first is the oscillator tuning or switching described above; the second consists in tuning the amplifier to resonance with the oscillator, which is indicated by a dip on the amplifier plate-current meter; and the third consists in coupling the antenna to the final amplifier and tuning it to resonance with the amplifier. Antenna resonance is indicated by maximum radiation on the antenna-circuit r-f ammeter. Each of these three adjustments must be repeated whenever a different frequency is used.

Marine radiotelegraph transmitters customarily employ fairly high-powered oscillators, with the purpose of obtaining maximum transmitter output while utilizing a minimum number of power amplifier stages. Oscillator tubes such as the UV211 and 860 are very commonly used, and power outputs from the oscillator of 100 w can be obtained with these tubes. However, such oscillators are usually run at much lower outputs with improved stability.

Fixed service (point-to point) radiotelegraph, broadcast radiotelephone, and other fixed types of radiotelephone transmitter utilize crystal oscillators. In such oscillators every effort is made to obtain high stability. Power output is of secondary importance, since space is not at a premium and sufficient amplifier stages may be used to produce any required output power. Usually small tubes are used, such as the 6SK7, 6L6, and so on. Such tubes operate at low voltages, making possible the use of small circuit constants, which permits economical design of the entire oscillator circuit with a high degree of stability. Many commercial broadcast transmitters employ tubes as large as the type 802. Although as much as 16 w may be obtained from an 802 oscillator, such tubes are customarily operated at much lower outputs in the interest of stability. In these oscillators the use of the electron-coupled circuit is virtually standard practice. In many transmitters, the oscillator plate circuit is untuned, which contributes materially to the reliability.

Where uninterrupted operation is important, as in broadcast transmitters, duplicate crystals and ovens are provided, with both ovens in continuous operation. In case of crystal failure a switching arrangement permits rapid changeover to the spare unit. Many manufacturers supply two complete oscillator units, a practice that has much to recommend it from the practical point of view, and the following buffer amplifier may be easily and rapidly switched to either oscillator. Each oscillator is constructed on a self-contained, shielded chassis: upon failure of one unit, the spare is switched in the circuit and the defective unit easily removed from the transmitter for service.

## THE TRANSMITTER AMPLIFIER

In general, transmitter amplifiers function to amplify the oscillator output to the value of r-f power that it is desired to present to the antenna system. In most modern marine radiotelegraph transmitters, considerable power is usually taken from the oscillator circuit, and the following transmitter amplifier therefore functions as a power amplifier. Most modern marine radiotelegraph transmitters utilize a single power-amplifier stage, employing a number of tubes in parallel to achieve the desired

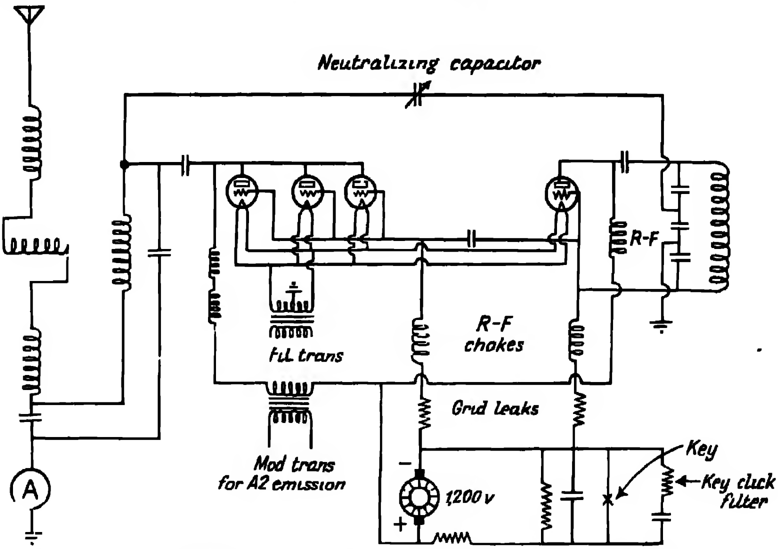



FIG. 189. Elementary master-oscillator power amplifier marine-type transmitter.

output. For this reason a fairly high powered oscillator is used, although at some sacrifice of oscillator stability. However, since such transmitters are not subject to continuous operation over long periods, and since the oscillator and amplifier circuits are frequently retuned, the sacrifice in stability is to some extent justified. The circuit of an elementary master-oscillator power amplifier marine radiotelegraph transmitter is shown in Fig. 189.

In broadcast transmitters, oscillator stability is the primary consideration. If appreciable power were drawn from the oscillator, instability of oscillator frequency, such as that caused by modulation of later amplifier stages, would be experienced under varying load conditions. To enable the oscillator to work into a constant load impedance under all conditions, an isolating amplifier is utilized between the oscillator and the power-amplifier section of the transmitter. An amplifier employed in this manner is called a **buffer amplifier.**

A buffer amplifier is essentially a voltage amplifier and is usually biased to at least plate-current cutoff point. For use with a given oscillator, a buffer amplifier tube is chosen having characteristics that will permit operation such that the positive peaks of oscillator r-f voltage do not exceed the necessary grid-bias value. The buffer therefore draws no grid current, and a nonvarying load is presented to the oscillator. Since considerable voltage gain is obtained in the buffer amplifier, relatively low power tubes can be utilized as oscillators with the circuit-stability advantages previously mentioned. Consequently, the buffer stage serves the double purpose of providing voltage amplification and isolating the oscillator.

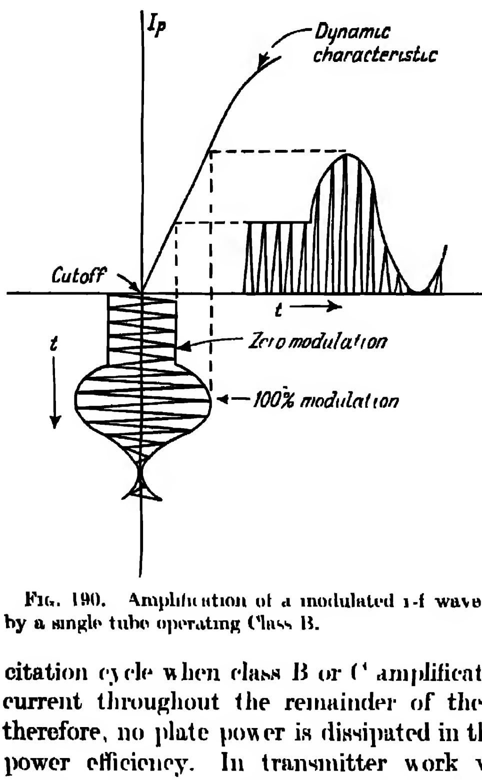



FIG. 190. Amplification of a modulated r-f wave by a single tube operating Class B.

Except for certain special-purpose stages, the amplifier stages following the buffer in a transmitter are power amplifiers. As power amplifiers, they may be operated class A, class B, or class C. Since classes B and C are relatively more efficient than class A amplification, these classifications are employed almost exclusively in transmitter r-f power amplifiers. Since plate current flows during only a portion of the excitation cycle when class B or C amplification is used, there is zero plate current throughout the remainder of the cycle. During this period, therefore, no plate power is dissipated in the tube, with resultant higher-power efficiency. In transmitter work where high-powered vacuum tubes are used, the power saving is appreciable.

With a-f class B amplification, it is necessary to use two tubes in push-pull to avoid distortion. Although the push-pull circuit is often used in r-f power amplifiers, a single-tube stage is often operated class B in transmitters without distortion of the modulated envelope. This is shown in Fig. 190. Here a single tube operating class B is used to amplify a modulated r-f signal. Despite the fact that the r-f carrier is considerably distorted by the negative peak cutoffs, the wave form of the a-f envelope is unaffected, provided that the tube is operating over a linear portion of its characteristic.

With single-tube class B operation, harmonics of the carrier frequency are present in the plate circuit. This necessitates the use of tuned circuits in the following amplifier stages to filter out the harmonics, which otherwise would reach the antenna and be radiated, causing objectionable interference on these frequencies. By the use of two tubes in push-pull operating class B, all the even-order harmonics are eliminated in the plate circuit if the stage is properly balanced, with the result that the amount of filtering required in subsequent stages is considerably reduced.

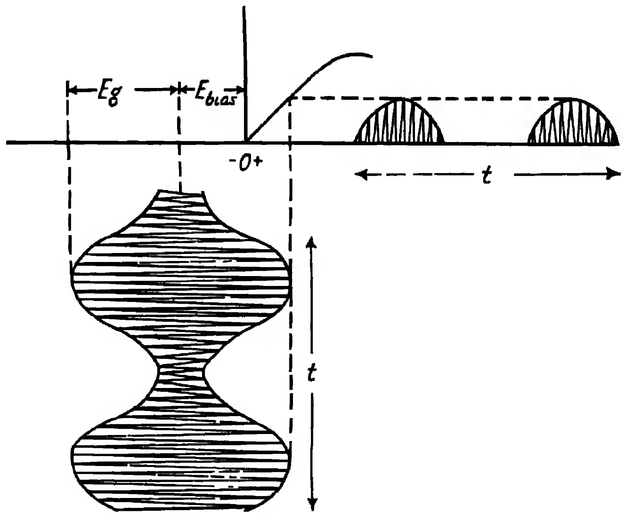



FIG. 191.—Distortion of modulated envelope caused by class C amplifier.

When a class B amplifier is utilized as a carrier frequency amplifier *before* the modulated stage of a radiotelephone transmitter, a plate-circuit efficiency of approximately 33 per cent can be realized. When a class B amplifier operates as an amplifier of a modulated r-f signal, the plate circuit efficiency is also a function of the percentage modulation. With 100 per cent modulation, the instantaneous peak power output is four times the unmodulated output. With the same percentage modulation, the average power output is 1.5 times the unmodulated output. Plate efficiencies of approximately 70 per cent can be realized with 100 per cent modulation.

Because of the higher efficiencies that are possible, class C r-f power amplification is used in transmitters wherever possible when considerable power is being handled. Because of the distortion of the modulated envelope (Fig. 191), class C amplification cannot be utilized to amplify a modulated carrier, but it is widely used in amplifier stages preceding the modulated stage in radiotelephone transmitters. In radiotelegraph

transmitters, where there is no modulation envelope to consider, the use of class C power amplification is practically universal, and plate-circuit efficiencies of approximately 85 per cent are realized under such conditions. The application of class B and class C amplification in the modulated stage of a transmitter is discussed in a later section of this chapter.

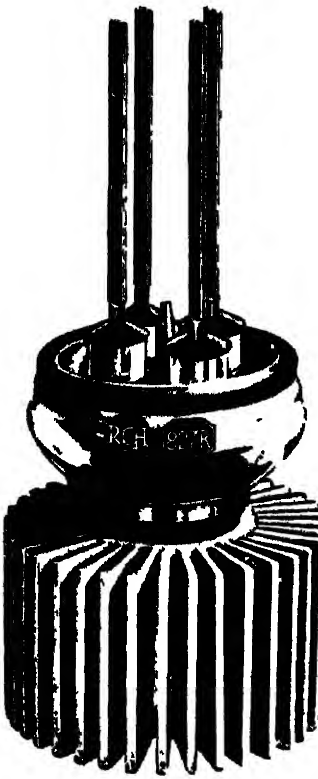




FIG. 192 An air cooled transmitting tube showing heat radiating fins (*Courtesy of RCA Manufacturing Co., Inc.*)

Power tubes are customarily rated by the manufacturers on the maximum power that can be dissipated at the plate without overheating. Thus, if an r-f power amplifier stage is designed to operate with an efficiency of 50 per cent, an input from the power supply of 100 w would be required to produce a desired output of 50 w. A tube that could dissipate 50 w (plate circuit) without excessive heating would be required. Increasing the plate circuit power input above the 100-w value in an attempt to obtain greater output would result in overloading the tube, with resultant overheating and possible destruction. By utilizing a different class of amplification, however, the plate efficiency can be increased, and the same tube can be used to achieve greater output. Thus, if the efficiency is increased to 70 per cent, the permissible input can be increased to 166 w, with a resultant output of 116 w, an increase of 232 per cent. The power dissipated in the tube remains the same, that is, 50 w. It is apparent that the higher the plate-circuit efficiency the greater the output which can be obtained from a given tube. It is for this reason that class C amplification, which results in highest plate-circuit efficiency, is employed whenever consistent with other circuit requirements and especially when high-power outputs are of primary consideration.

The power loss in a vacuum tube is dissipated in the form of heat. Although this loss can be mathematically treated as though it were an $I^2R$ loss in the tube internal plate cathode circuit, it is actually due to

electronic bombardment of the plate. The heat that is dissipated, therefore, is radiated from the plate. In small, receiving-type vacuum tubes, sufficient plate radiation of heat is obtained by allowing air to circulate freely about the glass envelope of the tube. The high heat-radiating efficiency of a black body is responsible for the development of plate structures composed of carbon or graphite. As the power ratings increase, the heat-radiation problem increases, and it is often necessary to increase air circulation about a tube by means of a blower. Some types of air-cooled tubes are designed to operate with the plate at a dull-red, or cherry-red, heat. Above 1,000-w power ratings, it is often necessary to resort to water cooling of the tube plate.

A typical high powered water-cooled transmitting tube is shown in Fig. 193. When operated as a class C plate-modulated r f power amplifier, the tube pictured an RCA type 898 can deliver an output of approximately 45 kw.

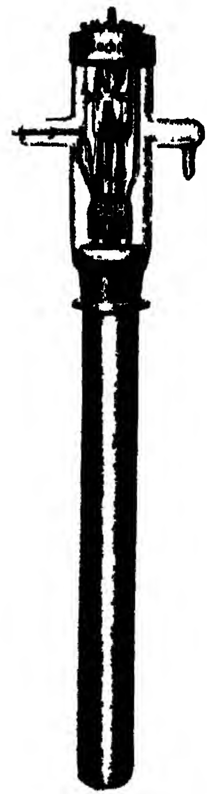


FIG. 193. A typical water-cooled transmitting tube. (*Courtesy of RCA Manufacturing Co., Inc.*)

The construction of a typical water-cooled power tube is shown in Fig. 194. The plate of the tube is a seamless copper tube, and by means of a special copper-glass sealing arrangement, this copper plate is welded to the glass base of the tube and forms part of the envelope. The filament and grid are supported inside the cylindrical copper anode and are not visible through the glass envelope. The copper-tube plate is immersed in water in a water jacket that completely surrounds it. Rubber hoses connect the water-jacket inlet and outlet to a pumping system that keeps the water circulating. Water with high specific resistance, usually rain water or distilled water, is used in order to prevent current leakage. Since the plate voltage on such tubes is on the order of 15,000 to 20,000 v, it is usually necessary to coil both inlet and outlet hoses in lengths of 20 to 30 ft to provide a long water-leakage path.

Most commercial vacuum-tube water-cooling systems utilize a closed circulating arrangement. Water pressure and resulting circulation are provided by an electrically driven centrifugal pump. A pressure valve

is so connected in the circuit that any decrease in water pressure acts to operate a relay, which, in turn, disconnects the plate voltage from the tube. Were it not for this precaution, the large amount of power dissipated at the plate would cause a rapid excessive rise in temperature upon failure of the cooling system and destroy the tube. Cases are on record where the copper anode of the tube has melted in such circumstances.

The exact circuit calculations involved in the design of class B and

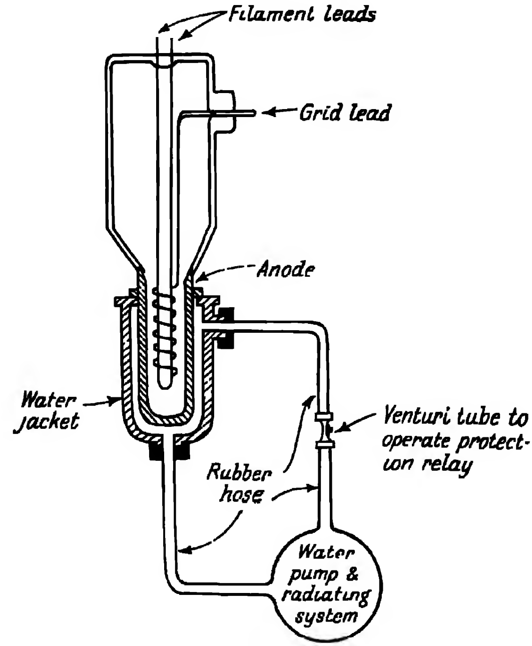



FIG. 194 Construction of a water cooled transmitting tube

class C r f power amplifiers become quite involved, mainly because of the number of variable factors concerned A precise method of design would entail making a number of different calculations to obtain the performance under various conditions. The ultimate design is a function of the degree of linearity desired, required power output, load impedance, and available excitation voltage. The procedure is to utilize the manufacturer's characteristic curves for the particular tube being considered and plot a number of graphs for different load impedances with variation of excitation voltage. The circuit constants are then chosen for the operating point at which the desired power output is obtained with the greatest degree of linearity.

An approximate procedure is to choose a tube that is capable of supplying the desired power output. The plate circuit efficiency, plate

voltage, and current under optimum load conditions are then obtained from manufacturer's data. With these values known, the necessary grid bias can be computed approximately from the amplification factor. Thus, class B amplification is defined as amplifier operation in which the grid bias is approximately equal to the cutoff value, so that plate current flows for approximately one half of each alternating voltage input cycle. Plate-current cutoff may be assumed to occur when

$$\mu E_g = E_p, \tag{17}$$

where $E_g$ = grid voltage;
$E_p$ = plate voltage.

Solving Eq. (17) for $E_g$,

$$E_g = \frac{E_p}{\mu}. \tag{18}$$

The approximate necessary value of negative grid bias to operate a given tube as a class B power amplifier may then be obtained by substitution in Eq. (18).

The calculation for class C amplifiers is based on the same method. Since plate current flows for considerably less than one half cycle with class C operation, the grid must be biased appreciably beyond cutoff. Class C operation varies considerably with particular applications depending upon other circuit requirements. Plate current may be required to flow during as small a portion of the cycle as 60° and during as large a portion as 120°. The bias value beyond cutoff is therefore a function of the type of class C operation desired. In theoretical problems where the exact type of operation is not specified, a bias value approximately 2.5 times the cutoff value is customarily assumed. The cutoff value is obtained from Eq. (18) and multiplied by the correction factor 2.5. Thus, for class C operation,

$$E_g = \frac{2.5E_p}{\mu}. \tag{19}$$

**Problem.** A triode transmitting tube operating with a plate voltage of 1,250 v has a filament current of 3.25 amp, a filament voltage of 10, and a plate current of 150 ma. The amplification factor is 25. What value of control grid bias must be used for operation as a class C amplifier?

**Solution.**

$$E_g = \frac{2.5E_p}{\mu}. \tag{19}$$

Substituting,

$$E_g = \frac{2.5(1,250)}{25}, \tag{20}$$

$$E_g = 125, \tag{21}$$

$$E_g = -125 \text{ v}. \tag{22}$$

**The Frequency Multiplier.** It is often desirable to operate a crystal oscillator at a comparatively low frequency to furnish excitation for an amplifier that operates at a much higher frequency. This is generally true of the operation of h-f transmitters. High-frequency oscillations are not readily produced by quartz-crystal oscillators. As has been shown (Chap. XI), the frequency of oscillation of a quartz crystal varies inversely with the thickness and crystals ground for high frequencies are therefore very thin, have a high natural period of vibration, and are very likely to crack in service. In addition, they are notoriously unstable, being much more subject to the effects of temperature variation.

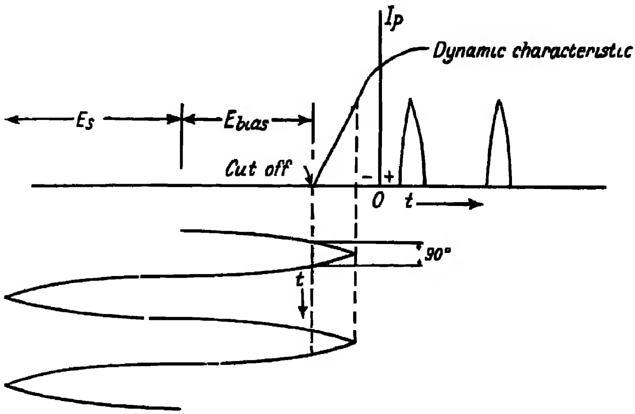
In modern h-f transmitters, the crystal oscillator is operated at a comparatively low frequency, which is readily produced by the crystal. High frequency output is achieved by means of *frequency-multiplier* stages. The plate current of a class C amplifier is badly distorted and therefore contains a large percentage of harmonics. It is possible to resonate the tank circuit of an amplifier to one of these harmonics and cause it to absorb considerable power at this frequency. The impedance offered to the fundamental and the remaining harmonics by the tank circuit will be small, and consequently, comparatively little power will be developed in the tank circuit at these frequencies. An amplifier operated in this manner is called a **frequency multiplier.**

The tank circuit of a frequency multiplier is usually tuned to the second harmonic of the input excitation frequency, since the second harmonic content of the amplifier plate current is ordinarily greater than that of the other harmonic frequencies. A frequency multiplier operated in this fashion is called a **frequency doubler.** Where very h-f output is required from a transmitter using an l-f crystal-oscillator stage, two or more doubler stages are often used. A class C amplifier having a plate efficiency of 80 per cent will show an efficiency of approximately 70 per cent when used as a frequency doubler.

Although the approved practice is to employ a number of frequency doubler stages to obtain a desired h-f output, greater multiplication per stage can be employed. Frequency tripling is often utilized where the number of stages which may be employed is limited. In transmitters of extremely high-frequency output, quadrupling of frequency has been used with relatively good results. However, the plate efficiency falls off rapidly as the multiplication per stage is increased with resultant decreasing power output. A number of compromise designs have been worked out by manufacturers. These designs utilize frequency tripling with an added stage of straight power amplification to make up the power loss in the multiplier stages. It has been possible to effect a design employing fewer stages than would be required if frequency doubling were utilized to obtain the same power output.

Since the input and output circuits of frequency multiplier stages are

tuned to different frequencies; troubles due to interstage coupling, feedback, and so on, are eliminated in such amplifiers. For this reason, triode tubes used as frequency multipliers need not be neutralized.



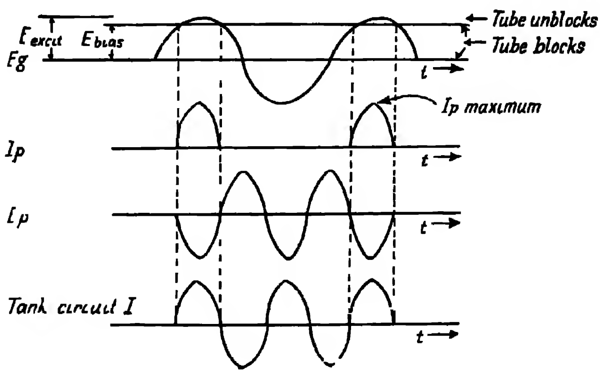




FIG. 195. (a) Class C amplifier operating as a frequency doubler. (b) Frequency-doubler wave forms showing relation between grid voltage, plate current, plate voltage, and tank current.

Frequency-doubling action is obtained in a class C amplifier by biasing the grid well beyond plate-current cutoff point, the necessary bias value usually being in the vicinity of twice that necessary for cutoff. This arrangement is shown in Fig. 195(a). At first it would appear that at such high values of negative bias the tube would be blocked throughout the entire input cycle. This is prevented, however, by increasing the excitation voltage sufficiently to cause plate current to flow during *one*

*fourth* of the excitation cycle. The harmonics generated by this mode of operation are called **forced harmonics.**

In Fig. 195(a). it is seen that plate current flows during only 90° of the grid-excitation cycle. Throughout the remaining 270° of the cycle, the tube is blocked by the excessive negative voltage on the grid. At the moment of *maximum* plate current, plate voltage is at a *minimum* owing to the *IZ* drop through the tank-circuit load. Since the tank circuit is tuned to parallel resonance at the second-harmonic frequency for frequency doubler operation, the impedance offered by it is at a maximum at this frequency. At zero plate-current condition (tube blocked), there is no *IZ* drop across the impedance of the tank circuit except that due to the circulating current, and plate voltage is consequently at a maximum. As the plate current increases, the *IZ* drop across the tank circuit increases, with resultant decrease in voltage at the plate of the tube. When plate current is at a maximum, therefore, the tank circuit *IZ* drop is maximum and plate voltage is at a minimum, as shown in Fig. 195(b). It is apparent that the plate voltage and current are 180° out of phase.

When plate current ceases during the 270° of the cycle in which the tube is blocked, the voltage continues to vary in an undistorted form around the average plate voltage because of the effect of the tuned-plate tank circuit. As power is transferred to the tank circuit because of the pulsation of plate current, energy is built up in the form of electrostatic and electromagnetic fields about the capacitance and inductance of this tank circuit. When the source of power is removed (when tube is blocked), the collapse of these fields induces an emf, which in turn causes a current in the tank circuit in the same form as the original pulses. This condition is called the tank circuit **flywheel effect** and is shown in Fig 195(b). The single pulse of plate current during 90° of the input cycle causes a complete cycle of plate voltage during 180° of the input cycle. It is apparent, therefore, that the plate voltage varies at twice the frequency of the excitation voltage, that is, there are two plate voltage cycles for each grid excitation-voltage cycle. Since the tank-circuit current depends directly upon the plate-voltage variations, the frequency of the tank-circuit circulating current is twice that of the excitation voltage; in other words, the input frequency has been doubled.

*Frequency tripling* is achieved by tuning the plate circuit to three times the exciting frequency. Better efficiency is obtained by permitting plate current to flow during only one sixth of the excitation-voltage cycle, which is accomplished by increasing the bias so that the tube unblocks during only 60° of the cycle. A complete plate-voltage cycle therefore occurs during 120° of the excitation-voltage cycle, that is, the tank-circuit circulating current frequency is three times that of the excitation voltage.

**Neutralization.** Since the advent of screen-grid tubes, with resultant reduction in tube interelectrode capacities, these tubes have been used almost exclusively in receiving circuits. Despite the fact that a number of screen-grid transmitting tubes are available in various power ratings, higher power efficiencies are possible in many applications with triode tubes, especially with the circuit parameters encountered in high-power installations. For this reason triode tubes are still employed in many commercial transmitter circuits. Many circuits employ a combination utilizing screen-grid tubes in some stages and triodes in others.

Because of the interelectrode capacitances of triodes, notably the plate-grid capacity, considerable feedback, or regeneration, normally occurs in such tubes through this capacitance in amplifiers where input and output circuits are tuned to the same frequency. Unless steps are taken to check such feedback, the amplifier breaks into self-oscillation. Even where feedback conditions are not severe enough to cause oscillation, the regenerative condition causes unstable operation and variation in frequency response and amplification.

Oscillation is prevented in transmitter power-amplifier stages by means of *neutralization*, which consists of feeding back from plate to grid circuit of the amplifier a voltage that is equal in magnitude but opposite in phase (180° out of phase) to the feedback voltage occurring through the tube interelectrode capacitance. This can be accomplished by inserting in the circuit a condenser having a capacitance equal to the tube grid-plate capacitance. The condenser couples the grid circuit to the proper portion of the plate circuit that provides a voltage having the necessary phase disposition. Neutralizing circuits must be carefully balanced to achieve the desired result. Any maladjustment may result in aggravating the feedback condition instead of relieving it. In addition, improper adjustment of neutralizing condensers may result in partial by-passing of the amplifier itself, with resultant loss in effective gain of the stage.

A number of circuits have been developed to achieve effective neutralization. Nearly all are variations of the Wheatstone-bridge circuit, in which the various bridge legs are capacitive or inductive reactances or combinations of both. Neutralization is accomplished when the bridge is balanced.

A number of commonly employed neutralizing circuits are shown in Fig. 196. The circuit shown at (a) is called the **Rice neutralizing circuit** after C. W. Rice, its originator. The bridge circuit equivalent circuit is also shown. In the equivalent-circuit diagram, the tube itself is omitted, the tube grid-plate capacitance being represented by $C_{gp}$. In transmitter circuits, the grid input inductance [$L_a$-$L_b$, Fig. 196(a)] is provided by the plate tank-circuit inductance of the previous stage. Accurate adjustment is facilitated by taps on this inductance. The lower terminal $N$ is connected to the plate by the neutralizing condenser $C_n$. The input terminals $G$ and $N$ and the output terminals $F$ and $P$ form the two pairs of diagonally

opposite points of the equivalent bridge circuit The bridge can be balanced in either sense Thus the condition of balance is attained when no voltage exists across terminals $GN$ because of a voltage across $FP$

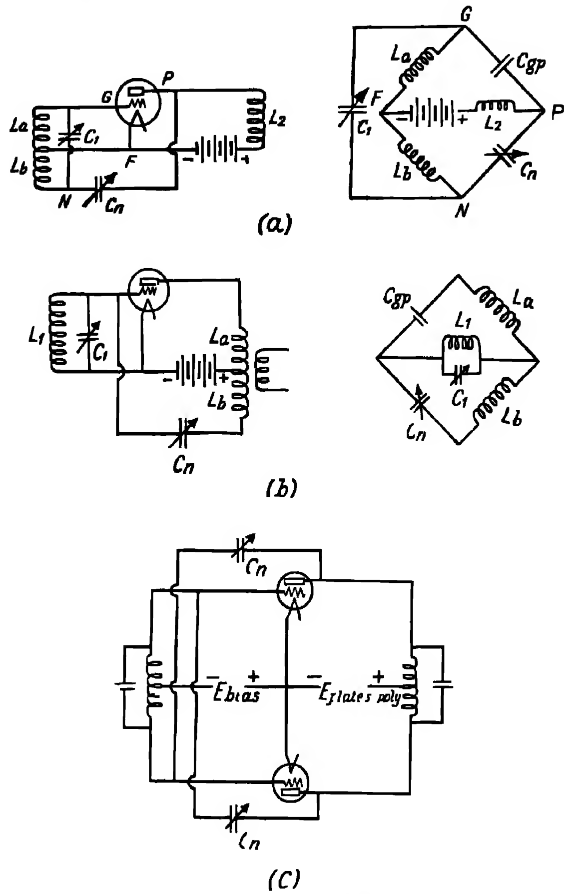



FIG 196 (a) Rice neutralizing circuit and equivalent circuit (b) Neutrodyne circuit and equivalent bridge (c) Neutralizing circuit for push pull amplifier

Similarly when the bridge is balanced because of a voltage across $GN$ no voltage will exist across $FP$ Hence no current will flow between $F$ and $P$ because of a voltage across $GN$ The latter method is the one commonly used in practice to achieve neutralization balance

The procedure is as follows The plate supply voltage is left off An r f thermocouple galvanometer is connected in series with the plate tank circuit inductance by means of a convenient switching arrangement Excitation is then applied to the grid, and if the circuit is unbalanced,

current will flow in the plate circuit and will be indicated by the galvanometer. Since no plate voltage has been applied to the tube, the current flow is not an electronic function. The current flow occurs through the tube capacity $C_{gp}$. Condenser $C'_n$ is then adjusted until zero current is indicated on the galvanometer. The stage is then neutralized, since the grid plate coupling through the tube has been completely counterbalanced by the out-of-phase coupling provided by the neutralizing condenser. Because all the bridge arms are not pure reactances, a resistance of approximately 25 ohms is customarily connected in series with neutralizing condenser $C'_n$. This arrangement permits a more exact phase balance to be obtained.

Another popular form of balance, or neutralizing circuit, known as the **neutrodyne circuit,** is shown in Fig. 196(b). This circuit was developed by L. A. Hazeltine and was very popular in the days of triode tuned r-f receivers. It can be seen that this circuit applies the bridge-balance arrangement in the plate circuit instead of in the grid circuit as the Rice system does, although the balance is obtained in much the same manner as in the Rice circuit.

A neutralizing circuit used in push-pull amplifiers is illustrated in Fig. 196(c). It can be seen that this arrangement is simply an application of the Rice circuit to each tube.

**Coupling Circuits.** The methods of coupling utilized between stages of a radio transmitter are, in general, similar to the coupling methods employed in receivers, namely capacitive coupling, inductive (transformer) coupling, impedance coupling, and so on. These coupling methods were discussed in detail in Chap. XII and will not be enlarged upon here. In transmitter circuits, provision is usually made to vary interstage coupling. This arrangement permits varying the excitation to an amplifier and also makes possible precise circuit adjustments.

Often in transmitters, succeeding amplifiers, or stages, are not in close physical proximity to each other, especially in high-power installations where succeeding power-amplifier stages are frequently housed in separate transmitter bays. In such cases, ordinary coupling methods between stages are not adequate.

This difficulty is overcome by utilizing *link coupling.* Here the output circuit of one amplifier is coupled to the input circuit of the following amplifier by a selection of low impedance transmission line of proper length. Appropriate impedance-matching transformers are employed at each termination of the line. A step-down transformer is used in the output circuit of the first amplifier to match the high plate impedance of the amplifier tubes to the low impedance of the line. At the other termination a step-up transformer is employed to match the low line impedance to the high amplifier input impedance.

In interstage coupling circuits at relatively low frequencies, the physical

distance is so small that the transmission-line dimensions do not approach the wave length of the circuit. The impedance of the line itself is consequently more or less a negligible factor Nevertheless, the low-impedance arrangement is utilized to keep losses in the line itself at a minimum.

In high-power transmitter installations, it is usually desirable to keep the transmitter building out of the more intense r-f field of the antenna in order to minimize absorption losses, feedback difficulties, and so on. The antenna is therefore located at some distance from the transmitter and coupled to it by a link circuit utilizing a long transmission line. In this case, the transmission line impedance is an important factor, and an exact impedance match must be established at both terminations if

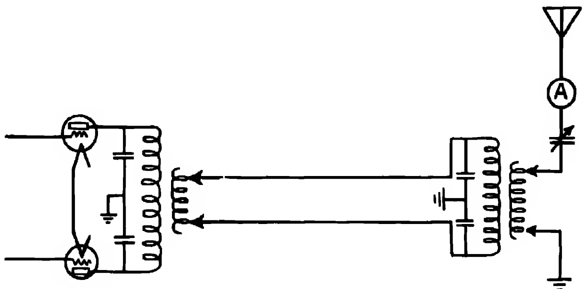



FIG. 197 A typical antenna coupling circuit

minimum power loss is to be attained. A typical antenna coupling circuit of this type is shown in Fig. 197.

## TELEGRAPH TRANSMITTER CONSIDERATIONS

Radiotelegraph transmitters are inherently less complex than radiotelephone transmitters. No voice modulation is employed, and the necessity for speech amplifier and modulator circuits is obviated. A radiotelegraph transmitter is essentially a device for generating an r-f carrier wave of a given power. Since there is no modulated-carrier envelope to consider, considerable distortion can be tolerated, the limiting factor being the amount of harmonic suppression required. For this reason, class C amplification finds wide application in all amplifier stages of a radiotelegraph transmitter with the possible exception of the buffer stage. It is therefore possible to achieve exceptionally high plate-circuit power efficiencies in such transmitters. If class C amplification is utilized throughout, the tank-circuit oscillatory wave train is completed by the flywheel effect previously discussed, and a substantially constant carrier frequency is supplied to the output.

Intelligence is transmitted by radiotelegraph transmitters by breaking up the r-f carrier into the dots and dashes of the international Morse

code. Since the radiated signal is unmodulated, it requires the use of a heterodyne-detector receiving system in order to become converted into an a-f signal. In this type of reception, a radio frequency is generated locally within the receiver and fed simultaneously with the transmitter signal into the receiver detector circuit. The local r-f signal differs by an audio frequency from the incoming telegraphic carrier. The resulting heterodyne therefore occurs at an audio frequency that is amplified and converted into sound in the usual manner.

It is often desirable to be able to receive telegraphic signals without the aid of a heterodyne-type receiver. To facilitate such reception, radiotelegraph transmitters are sometimes modulated at some fixed audio frequency, usually 500 or 1,000 c. Such modulation is customarily accomplished by utilizing a 500-c alternating current (provided by a special motor generator) as a source of plate voltage. Another popular and effective method is to interrupt the carrier at an a-f rate by some mechanical means, usually by a motor-driven segmented disk similar to a motor commutator. Such a device is called a **chopper.** The chopper interruptions should not be confused with the manual interruption of the r-f carrier into code dots and dashes by a telegraph key; for the chopper interruptions occur in addition to the telegraphic interruptions. This type of signal is known as **interrupted continuous wave** (sometimes abbreviated i-c-w). A telegraph transmitter producing a carrier that is unmodulated is said to produce a **continuous-wave** (c-w) signal.

Most modern radiotelegraph transmitters provide a switching arrangement by means of which either a c-w or i-c-w signal may be transmitted. The i-c-w transmission is often employed when it is desired to broaden the signal. The broadening effect occurs by virtue of the side bands produced by the modulation.

**Keying Systems.** As previously stated, telegraphic transmission is accomplished by breaking up the r-f carrier into the dots and dashes of the international Morse code. This action amounts to starting and stopping the transmitters in accordance with the configuration of the code characters. There are a number of ways in which keying may be accomplished.

One of the simplest methods of keying is to insert the key, or keying relay, in series with the primary of the power transformers. When the key is up, or open, power is removed from all circuits of the transmitter. This method is used for the small low-power transmitters utilized as marine emergency transmitters. It is unsuitable for transmitters of higher power because of the large amount of current being interrupted.

One of the most widely used keying systems is the so-called **grid-blocking system.** This arrangement, shown in its elementary form in Fig. 198(a), consists of applying excessive negative bias to the tube when the key is open. The high value of bias drives the tube to cutoff.

When the key is depressed, a section of the voltage-divider resistor is shorted by the key, removing the bias and permitting the tube to function normally.

In marine medium-power radiotelegraph transmitters, the grid-blocking system is applied to the oscillator tube, the excessive bias in key-up position preventing the tube from oscillating. In some circuits, the grid blocking circuit is extended to all tubes, both oscillator and amplifier. The circuit is merely an extension of that shown in Fig. 198(a).

Another grid-blocking system acts to key a transmitter by alternately inserting and removing an r-f choke coil from the grid circuit of the

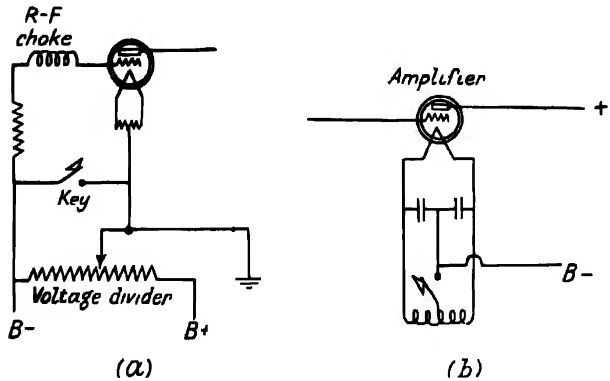



FIG. 198 Typical telegraph transmitter keying circuits (a) Grid blocking keying circuit (b) Center tap keying circuit

oscillator circuit. The presence of the choke effectively prevents the tube from oscillating When the key is depressed, the choke coil is short circuited and thus removed from the circuit permitting the tube to oscillate freely.

In crystal-controlled transmitters, oscillator keying is not feasible, because of the unstable circuit conditions that result and the tendency of the crystal to stop oscillating. In such transmitters, the grid-blocking method is applied to one of the power amplifier stages following the buffer. In high powered transmitters, keying is usually achieved in one of the low level stages.

Another system of keying, often called the **center-tap method,** is shown in Fig. 198(b). In this circuit, keying is accomplished in the negative or return, high-voltage circuit. The key, or relay, contacts are inserted in series with the negative high voltage terminal and the filament-transformer center-tap connection of one of the amplifier tubes. This system has the advantage that the plate current to only one tube is interrupted by the key.

A properly designed keying system produces clean-cut dots and dashes

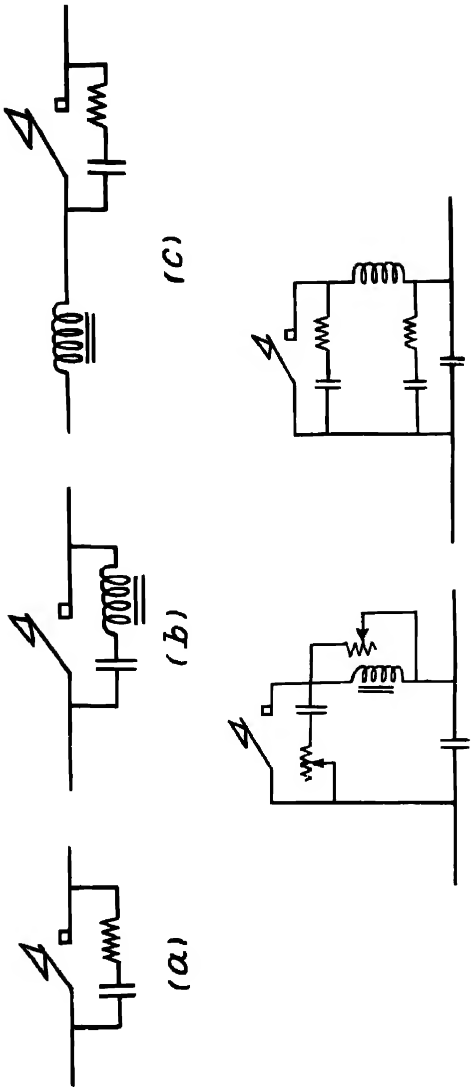



FIG. 190 A number of lag filter circuits for eliminating key clicks.

having a constant carrier frequency and producing a minimum of interference. When power is applied and removed from a circuit very suddenly, as in a keying circuit, the energy thus instantaneously released surges back and forth in the circuit until a condition of equilibrium is reached. This oscillatory condition is sharply damped in most keying circuits and therefore has no sharply defined natural period. Such an oscillation can be radiated from the keying circuit and can be picked up by receivers operating on widely different frequencies from that of the telegraph transmitter. Because of its short duration, this interference is heard as a *click* or *thump* in the receiver. Within a few kilocycles of the transmitter frequency, key clicks may also be radiated and cause interference to other transmitters operating in adjacent channels many hundreds of miles away.

There are two general methods of eliminating key clicks. The first is to use a time-delay filter circuit in association with the keying so that the transmitted signals are turned on and off gradually rather than abruptly, thus preventing the generation of high-order side bands. The circuits used for this purpose are appropriately termed **lag circuits,** and a number of representative lag circuits are shown in Fig. 199. In many keying circuits, the major click component occurs when the circuit is opened, rather than when the circuit is closed In such circuits, the lag circuit of Fig. 199(a) can be used successfully. When the key is opened, the circuit energy is released through the condenser instead of discharging in the form of an arc across the key contacts. Upon closing the key again, the condenser energy tends to dissipate across the key contacts, again causing a spark. Dissipation of energy is prevented by the insertion of a small resistor in series with the condenser which absorbs most of this energy.

The remaining lag circuits are seen to be various combinations of inductance and capacitance. Since an inductance possesses the property of opposing a sudden change in current, an inductance in series with the key causes the current to build up at a slow rate, the excess energy being stored up in the form of an electromagnetic field. When the key is opened, the energy stored in the electromagnetic field of the inductance is suddenly returned to the circuit. If the current is very large, there will be a tendency for an arc to form across the key contacts with an accompanying click. This secondary click is prevented by the action of the capacitance across the contacts as previously explained.

The second method of key-click elimination consists of inserting an r-f filter in the keying circuit. This circuit acts to filter out the keying transient before it can reach a part of the circuit from which radiation is possible. A typical r f absorption filter of this kind is shown in Fig. 200, in which the circuit consists of a one- or two-section inductance-capacitance filter arrangement. Many commercial manufacturers utilize

a combination of both a lag circuit and an r-f absorption filter. Regardless of whether or not an r-f filter is used, however, the lag circuit is almost always used, since, in addition to eliminating key clicks, the lag circuit prevents undue wear of key or relay contacts due to sparking.

Substantially all commercial radiotelegraph transmitters utilize keying relays for the actual interruption of the circuit. This practice permits low voltages to be used at the key itself and also eliminates undesirable long leads, since the key is usually located at an operating table more or less removed from the transmitter itself. In addition, the use of a relay facilitates the performance of other circuit operations simultaneously with the keying. Thus, in marine installations, the use of special double-contacted relays permits *break-in* communication. Normally, without the break-in relay, it is necessary to disconnect the antenna from the receiver during transmission in order to prevent the transmitter signal from blocking the receiver and possibly causing more serious disability. During transmission, therefore, the receiver is inoperative, and the operator cannot be "broken" or interrupted by the operator at the distant station to whom he is sending, and it is necessary to wait until the end of the entire transmission before asking for repeats and so on. When conditions are poor and the transmission is long much confusion can result from this procedure.

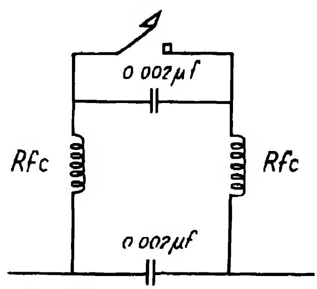



FIG. 200 An r f absorption filter circuit to prevent key click radiation

When a break-in relay is used, the sending operator can listen in on his receiver during transmission and thus can be interrupted at any point. The break-in relay employs an extra set of contacts that are connected in series between the antenna and the receiver. When the key is depressed and the transmitter radiates, the same relay armature movement simultaneously opens the receiver antenna circuit, disconnecting the receiver from the antenna. When the key is opened, the antenna is reconnected to the receiver. Consequently, the receiver is connected to the antenna between the dots and dashes of the transmission, permitting essentially continuous reception during transmission. In most marine installations, an additional set of contacts also breaks the circuit between the antenna and the transmitter when the key is up, thus preventing any detuning effect on the receiver. The transmitter-antenna contacts are so adjusted that they close just before the keying contacts close and open just after the keying contacts open. This lag in operation prevents arcing across these contacts, and actually the contacts are "dead" during opening or closing operating.

In many point to point shore radiotelegraph stations, automatic telegraph transmission is utilized. A machine with a typewriter keyboard is used to perforate an oiled paper tape in accordance with a code arrangement depending upon the type of operation. The tape is mechanically fed through a photoelectric cell or mechanically actuated keying device that is usually called the **transmitter.** The transmitter output, which is in the form of the dots and dashes of the international Morse code, is utilized to operate a relay, which in turn operates the radio transmitter. In such systems telegraphic speeds up to 250 words per minute can be achieved. At such speeds electromechanical relays are inefficient from

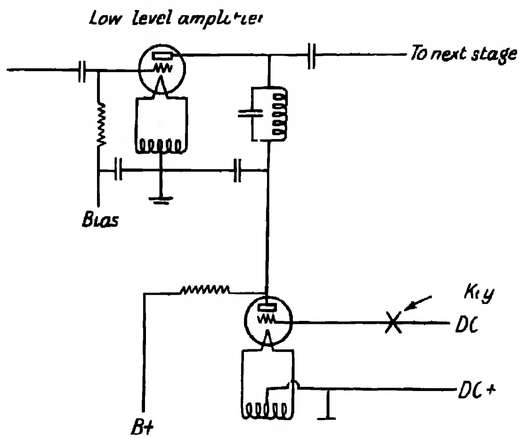
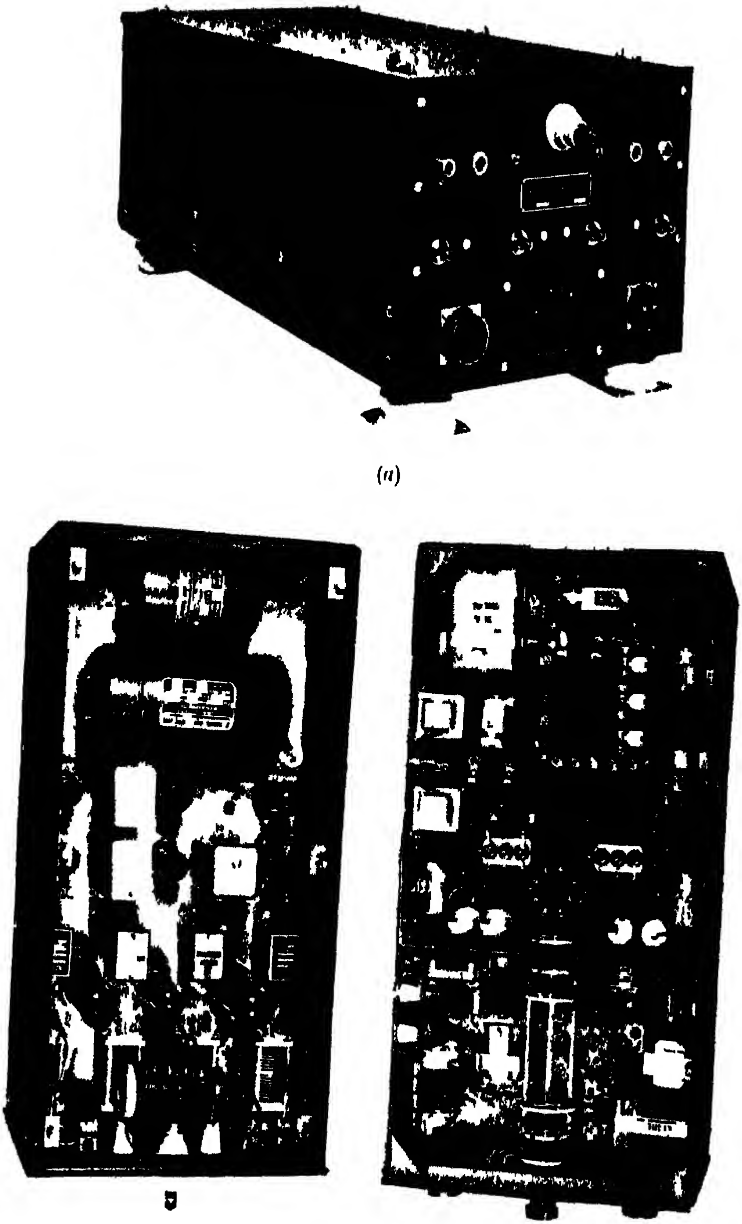



Fig. Electronic transmitter keying system.

the point of view of maintenance and adjustment. Electronic keying is therefore employed to replace the relay in such circuits. The d.c. keying output is employed alternately to supply and to remove a blocking bias from the grid of a vacuum tube called the **keying tube.** The plate of the keying tube is fed in parallel with one of the low level amplifier stages of the transmitter through a common resistor. When the key is up, there is no bias on the keying tube and a large plate current is drawn through the common resistor. Because of the resulting voltage drop across the resistor, the potential applied to the transmitter amplifier plate is reduced to a very low value and this stage does not supply sufficient output to excite the following amplifier stage. When the key is down, a negative bias beyond cutoff value is applied to the keying tube and no plate

Fig. 202. Composite radiotelephone unit. (a) Front view. (b) Top view with cover removed. Note dynamotors to provide receiver and transmitter plate power. (c) Bottom view with shield removed. The entire unit weighs 36 lb. It includes a pretuned 20 w four frequency crystal controlled transmitter and a four frequency crystal controlled superheterodyne receiver. (*Courtesy of Communications Co., Inc.*)



(*a*)  (*b*)  (*c*)

FIG. 202. (*See opposite page for caption.*)

current flows through this tube. Consequently there is no voltage drop due to this tube through the resistor and the amplifier functions normally, providing full excitation to the succeeding stage. The circuit arrangement is shown in Fig. 201.

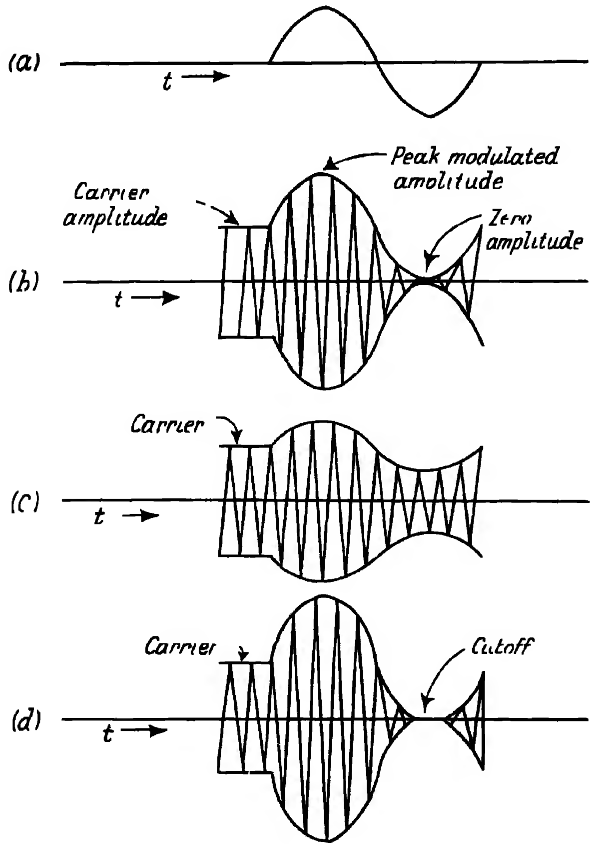



FIG. 205. Wave forms for various degrees of modulation. (a) Modulating a-f voltage. (b) 100 per cent modulation. (c) Undermodulation, less than 100 per cent. (d) Overmodulation, greater than 100 per cent.

## TELEPHONE TRANSMITTER CONSIDERATIONS

Radiotelephone transmitters are in general considerably more complex than radiotelegraph transmitters. In addition to generating an r-f carrier wave of the required power, the characteristics of a complex a-f current must be superimposed upon this carrier with a high degree of fidelity.

**Modulation.** The type of modulation in general use in the majority of radiotelephone transmitters today is called **amplitude modulation.**

*Unless specifically stated otherwise, the use of the word "modulation" throughout this chapter will refer to amplitude modulation* In radiotelephone transmitters for voice transmission, such as broadcast transmitters, amplitude modulation may be defined as the process by which the amplitude of the transmitted r-f carrier wave is varied in accordance with the sound waves actuating the microphone. These sound waves are converted into a-f electric currents, as explained in Chap. XIV. By successive stages of a-f amplification, the audio frequencies are amplified until sufficient power is available to transpose them effectively to the r-f carrier wave. Modulation can therefore be more narrowly defined as superimposing an a f current upon an r f current in such a manner that the r f current varies in amplitude in strict accordance with the variations in amplitude of the a f current. In actual voice transmission, a great number of different a-f currents are simultaneously imposed upon the r-f carrier so that the resulting wave form becomes extremely complex.

The degree of modulation occurring in a radiotelephone system is described in terms of the amplitude variation of the carrier wave. This is customarily expressed as a decimal value called the **modulation factor** or as a percentage called the **modulation percentage.** Various degrees of modulation are depicted graphically in Fig. 203. The modulation factor may be obtained from the equation

$$m = \frac{E_{max} - E_{min}}{2E_{ave}} \tag{23}$$

where $m$ = modulation factor;
$E_{max}$ = maximum carrier-voltage amplitude;
$E_{min}$ = minimum carrier voltage amplitude;
$E_{ave}$ = average carrier-voltage amplitude (amplitude of unmodulated carrier voltage).

The percentage of modulation is obtained by multiplying Eq. (23) by 100; thus,

$$M = \frac{100(E_{max} - E_{min})}{2E_{ave}}, \tag{24}$$

where $M$ = percentage modulation.

The maximum permissible modulation percentage, if distortion is to be avoided, is 100 per cent (see Fig. 203). Therefore, the maximum permissible modulation factor is 1. Any lesser degree of modulation will be represented by a modulation factor having a value between 0 and 1 and will be a decimal value. The method of obtaining the voltage amplitudes for substitution in Eqs. (23) and (24) is described in Chap. XVII.

**Plate-Circuit Modulation.** The most widely used type of modulation system is that in which the modulating signal is applied to the plate circuit of a power amplifier. In any modulation system, the circuit

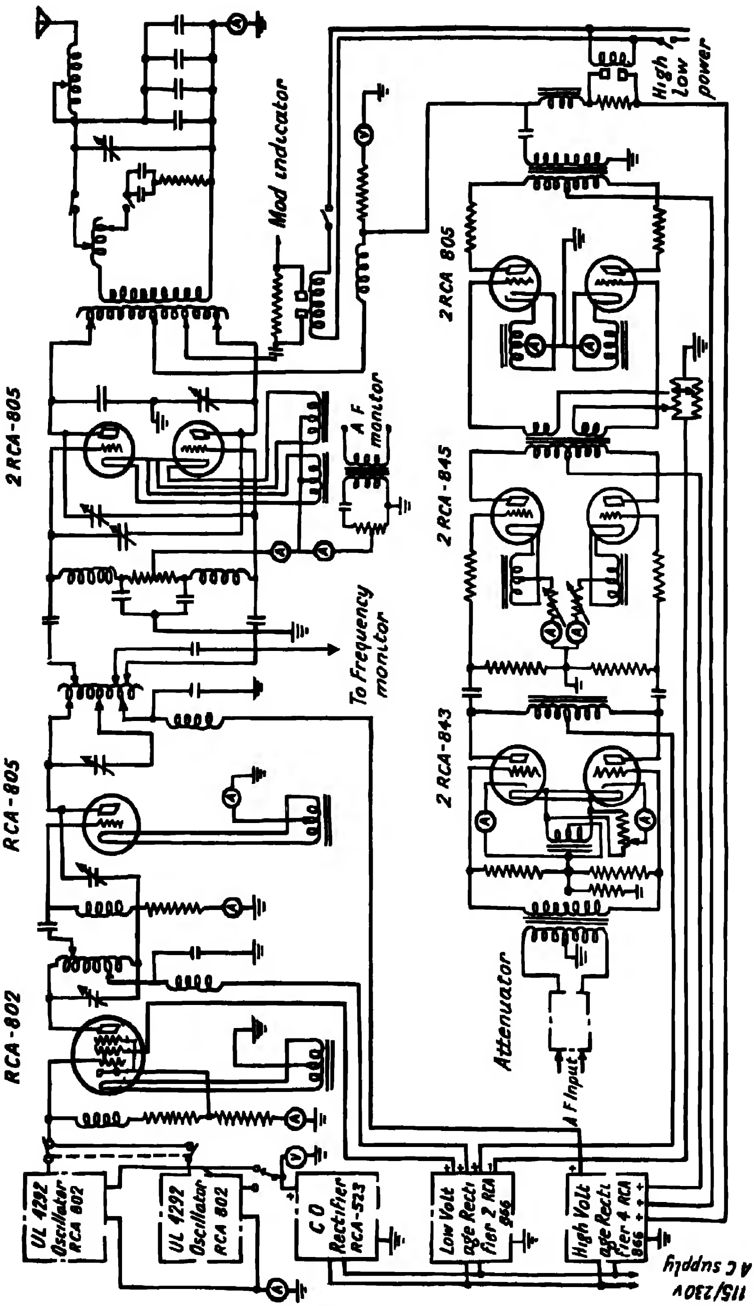



FIG. 204. Typical low power (100 to 250 w.) broadcast transmitter employing high level plate modulation (Courtesy of RCA Manufacturing Co., Inc.)

supplying the modulating power, that is, the final stage of a-f power amplification, is called the **modulator,** or **modulating stage.** The r-f power amplifier stage of the transmitter to which this a-f power is being supplied is called the **modulated stage.** In plate circuit modulation systems, the

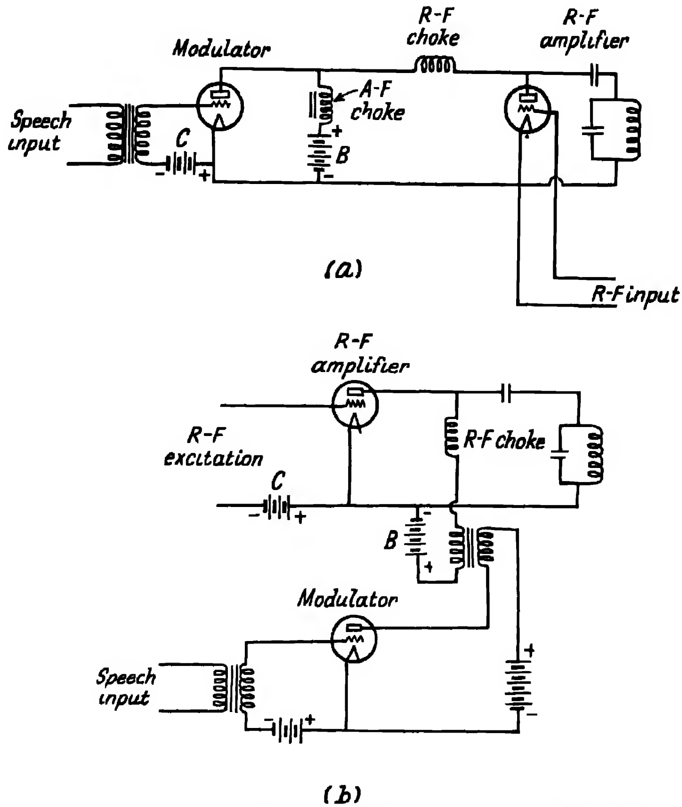



Fig. 205 Heising modulation circuits (a) Choke coupled (b) Transformer coupled.

plate circuit of the modulator is coupled to the plate circuit of the modulated stage. There are two basic methods in general use for accomplishing plate modulation: one is called **Heising,** or **constant-current,** modulation, and the other is called **modified Heising,** or **transformer-coupled,** plate modulation.

A typical Heising modulation system is shown in Fig. 205(a). Both the modulator and the modulated tubes are fed in parallel from the same power supply through a common a-f choke. The inductance of the choke tends to oppose any change in the current flowing through it. An *increase* in modulator plate current due to excitation from the preceding speech

amplifier causes a *decrease* in current to the plate of the modulated tube. The a-f choke holds the current through it from the power supply at a constant value; hence, the name **constant-current modulation.** Similarly a *decrease* in modulator plate current causes an *increase* in modulated tube plate current.

A transformer-coupled plate modulation system is shown in Fig. 205(b). The amplified speech or music (audio frequencies) in the modulator plate circuit is delivered to the output or load transformer, the secondary of which is connected in series with the modulated-stage plate supply. This transformer is commonly called the **modulation transformer.** The system is so designed that for modulation of 100 per cent the a-f voltage developed across the secondary of the modulation transformer is exactly equal to the plate voltage of the modulated tube when the a-f excitation to the modulator is at maximum. At positive modulation peaks, the modulated-stage plate voltage and, hence, current are doubled. Conversely, at negative modulation peaks, the plate current drops to zero with 100 per cent modulation. Thus, the modulated-stage plate current, which already contains an r-f carrier component, is also caused to vary in accordance with a-f modulating signals. Since any transformer is also an impedance-changing device, impedance matching must be taken into consideration in transformer coupled plate-modulation systems if minimum distortion is to be incurred. The transformer-turns ratio should be so chosen as to match properly the plate impedance of the modulator to the load impedance, that is, the plate impedance of the modulated stage. The reader is referred to Chap. X for a discussion of optimum impedance matching with minimum distortion.

In general, modulation is applied to the transmitter in one of two different ways: by *high-level* modulation or *low-level* modulation. In low-level systems, the modulating power is applied to one of the low-power-level amplifiers of the transmitter, that is, there are one or more linear power amplifiers following the modulated stage. In high-level systems, the modulating power is applied to the final power-amplifier stage. High-level modulation requires the development of considerable a-f power from the modulator. However, this system permits the use of high-efficiency class C amplifiers in the stages preceding the final stage, since distortion is not a controlling factor in these amplifiers. When low-level modulation is employed, linearly operated class B amplifiers with resultant lower plate efficiencies must be used in the stages following the modulated stage if distortion is to be avoided. The advantage of low-level systems is that relatively smaller amounts of a-f modulating power are required. Most modern commercial transmitters effect a compromise design in which low-level modulation is applied to one of the intermediate, medium-power amplifiers.

The power that a modulator stage must be capable of supplying for

100 per cent modulation depends upon a number of factors. With 100 per cent modulation, at the *peak* of the modulated signal the amplitude

FIG. 206. 5,000-w broadcast transmitter. (a) Front views with doors open. Class B modulator stage to the left. Class C power amplifier to the right. (b) Rear views with doors open. Note motor-driven blowers for air-cooling the high-power vacuum tubes. (*Courtesy of Collins Radio Co.*)

(either voltage or current) is doubled The *instantaneous peak power* will consequently be four times the unmodulated power as can be seen from an inspection of the familiar power equation

$$P = I^2R \qquad (25)$$

It is apparent that if the value of $I$ is doubled the value of $I^2$ and hence, of $P$ is quadrupled

With continuous 100 per cent modulation by a single pure audio tone, the *average* power will be 1 5 times the unmodulated power The average power when *speech modulation* is used which has the same instantaneous peak amplitude as a single pure tone has been shown to be approximately half as great as the average power in tone Therefore, despite the fact that the instantaneous peak power of a speech modulated wave is four times the carrier power just as with tone modulation the average power is only 1 25 times the unmodulated power Because the average power of a speech or music modulated wave varies greatly under actual conditions, the single tone value of average power (1 5 times unmodulated power) is customarily used as a basis for circuit calculations Since the current in a circuit is proportional to the square root of the power ($I = \sqrt{P/R}$) the circuit current at 100 per cent modulation will be 22 5 per cent greater than with carrier alone

If an unmodulated power amplifier required a power input of 100 w a power input of 1 5 times 100 w or 150 w would be required to modulate this stage 100 per cent In other words the modulator would be required to supply the additional 50 w in the form of a f power This additional 50 w of power is expended in the radiated signal in the form of *side bands* Side bands are the result of the heterodyning of the carrier frequency and the modulating audio frequency There are two side bands corresponding to the sum and difference heterodyning frequencies The upper side band is equal to the carrier plus the audio frequency and the lower side band is equal to the carrier frequency minus the audio frequency There will be a pair of side bands for each separate audio frequency in the modulating signal The frequency band occupied by the transmission will therefore be equal to twice the highest modulation frequency If the wave form is distorted by overmodulation the resultant harmonics will create additional side bands with consequent broadening of the transmission band For this reason overmodulation is prohibited by the FCC

**Problem.** A modulated power amplifier, operating at 50 per cent efficiency, is delivering an output power of 270 w with 100 per cent modulation If the modulator is operating at 60 per cent efficiency what is the power input to the modulator?

**Solution.**

$$\text{efficiency} = \frac{\text{output}}{\text{input}} \qquad (26)$$

Substituting for the power amplifier (modulated) stage,

$$0.5 = \frac{270 \text{ (watts output)}}{\text{input (watts)}}, \tag{27}$$

and

$$\text{input} = \frac{270}{0.5} = 540 \text{ w.} \tag{28}$$

Assuming the theoretical condition of modulation of a single pure tone as specified above, 540 w is 1.5 times the unmodulated carrier power. The unmodulated carrier power input is therefore two thirds of 540 w, or 360 w, and the power supplied by the modulator is the remaining one third, or 180 w.

Since the modulator is operating at 60 per cent efficiency, it is apparent that

$$\text{power input} = \frac{\text{power output}}{\text{efficiency}}, \tag{29}$$

or

$$\text{power input} = \frac{180 \text{ w}}{0.60} = 300 \text{ w.} \tag{30}$$

**Grid Modulation.** In a grid-modulation system, often called **grid-bias modulation,** the secondary of the modulator output transformer is connected in series with the grid-bias supply of the amplifier to be modulated. A typical grid-bias modulation system is shown in Fig. 207. When the grid bias, r-f excitation and load circuit of the modulated amplifier are properly adjusted, 100 per cent modulation can be obtained with fairly high efficiency. However, the plate efficiency of the modulated stage is considerably lower than with plate modulation. This system, in effect, amounts to varying the grid bias at an a-f rate.

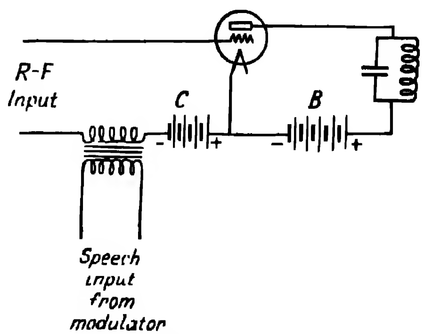



FIG. 207. Grid-bias modulation system.

The major advantage of grid modulation is that very little a-f energy is required for complete modulation. The modulator is required to furnish only a portion of the grid excitation losses of the modulated amplifier. The power in the side bands is obtained from the plate circuit. Thus, the size and expense of the a-f modulator and speech-amplifier equipment are greatly reduced.

Grid-modulated amplifiers are operated as class C amplifiers. A linear dynamic $E_g$-$I_p$ characteristic is ensured by making the plate tank-circuit impedance larger than is normally required. The r-f excitation voltage, a-f modulating voltage, and grid-bias voltage are connected in series.

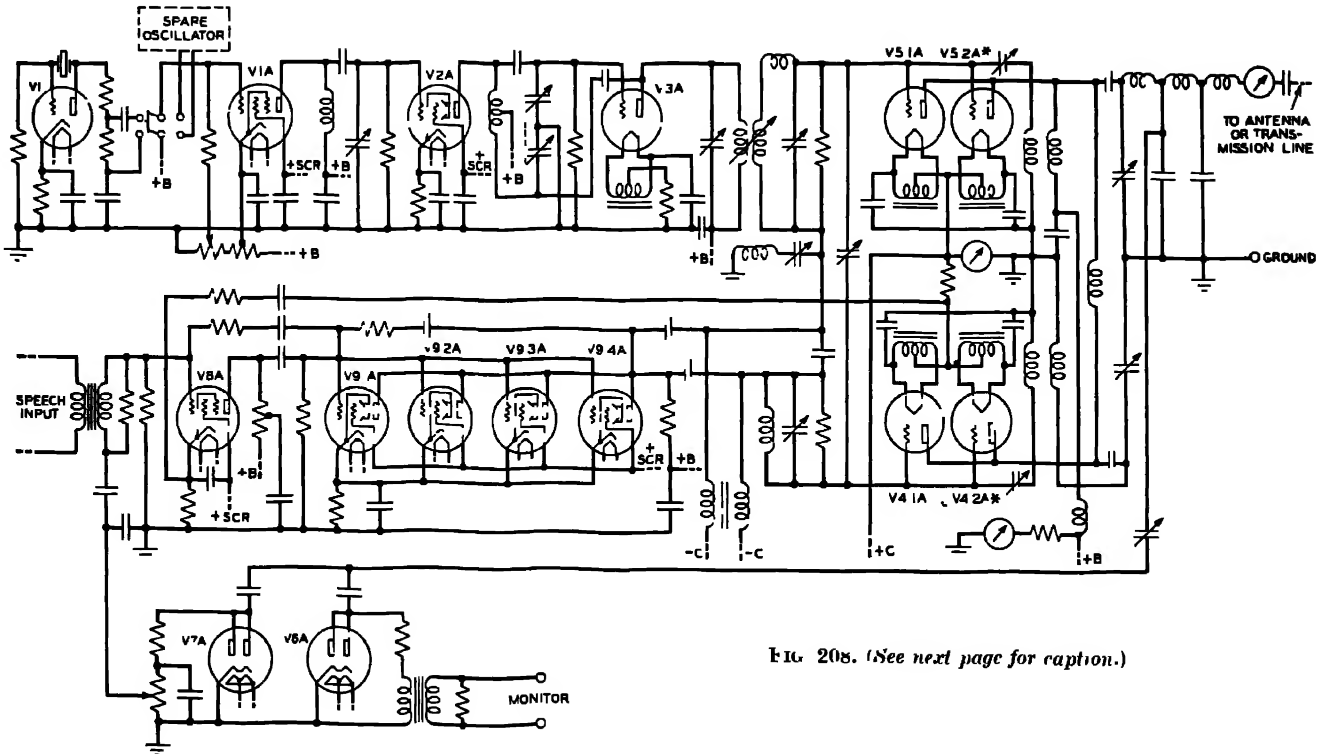



FIG. 208. (See next page for caption.)

Consequently, the modulating voltage alternately adds to, or subtracts from, the grid bias, causing a corresponding variation in the plate-current pulses. On modulation peaks, with 100 per cent modulation, the stage operates as a class B amplifier. When unmodulated, the mode of operation changes to that of an underexcited class C amplifier. Because of the low plate-circuit efficiency and the difficulty of securing linear operation, grid modulation is not so widely used as would be expected from its economy.

Most plate-modulated amplifiers are customarily operated as class C amplifiers. Despite the fact that a modulated signal cannot be amplified by a class C amplifier without distortion, as shown in Fig. 191, a class C amplifier can itself be modulated without distortion. This is due to the fact that the mode of operation varies from class C to class B when a modulating signal is present. The modulating voltage is superimposed upon the plate voltage, and the plate voltage therefore rises and falls with the modulating frequency. This amounts to shifting the operating point of the tube along the $E_g$ $I_p$ dynamic characteristic, with the result that the mode of operation varies between class B and class C, depending upon the amount of modulation. The advantage lies in the fact that with low levels of modulation the power consumed is small with resultant high plate circuit power efficiency.

The same is true for grid-modulated class C amplifiers. Class C grid modulation is depicted graphically in Fig. 209. The difference between the excitation wave form shown and that of a modulated carrier frequency should be noted. The modulating a-f voltage has the effect of varying the effective grid bias, which amounts to varying the operating point

FIG. 208 Western Electric 1,000 w broadcast transmitter The Doherty linear final amplifier is grid bias modulated and operated at a plate efficiency of approximately 60 per cent Tubes are identified in the list below (*Courtesy of Western Electric Co, Inc*)

| *Designation* | *Code Number of Tube* | *Designation* | *Code Number of Tube* |
|---|---|---|---|
| V1 | W.E 247A | V6A | W.E 351A |
| V1A | W.E 349A | V7A | W.E 351A |
| V2A | W.E 350A | V8A | W.E 349A |
| V3A | W.E 331A | V9.1A | W.E 350A |
| V4.1A | W.E 357A | V9.2A | W.E. 350A |
| V4.2A* | W.E 357A* | V9.3A | W.E. 350A |
| V5.1A | W.E. 357A | V9.4A | W.E 350A |
| V5.2A* | W.E 357A* | | |

The following tubes are used rectifier (not shown)

| | | | |
|---|---|---|---|
| V10A | W.E 301A | V13A | W.E 249B |
| V11A | W.E 267B | V14A | W.E 249B |
| V12A | W.E 267B | | |

* Tubes V4.2A and V5.2A are not provided for the 442A 1 radio transmitting equipment.

of the tube along the dynamic $E_g$-$I_p$ characteristic at an a f rate Despite the fact that the plate circuit efficiency of grid modulated amplifiers is low in comparison with other types of modulation, the plate efficiency of a grid modulated class C amplifier is considerably higher than any other class of amplifier similarly modulated

**The Doherty High-efficiency Amplifier.** The major function of a transmitter power amplifier is to develop as much power output as

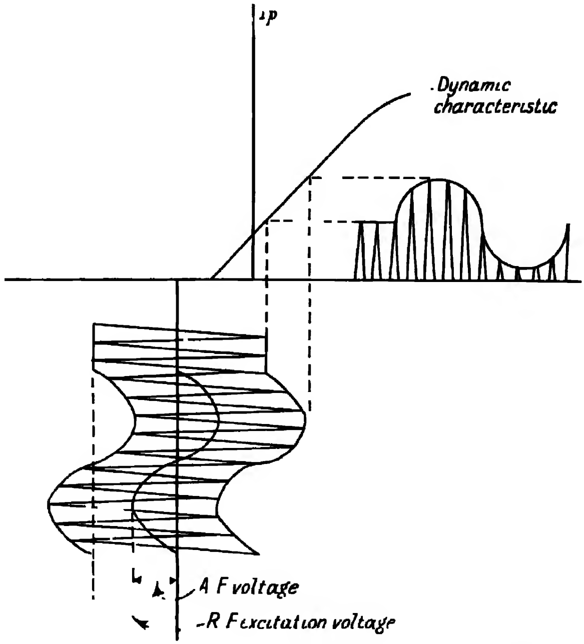



Fig 209 Wave form of a grid modulated class C amplifier

possible In amplifiers following the modulated stage where distortionless amplification is required linear operation of the amplifier is obtained at the sacrifice of considerable plate power efficiency A special linear amplifier developed by W H Doherty is shown in Fig 210 Efficiencies approaching those of class C unmodulated amplifiers are possible with this circuit with linear operation The Doherty high efficiency linear amplifier has recently had wide application in modern broadcast transmitters and is often used as a grid modulated amplifier

The Doherty amplifier utilizes two tubes Tube $T_1$ (Fig 210) delivers its output to the load through an artificial line equivalent to a quarter wave transmission line Tube $T_2$ delivers its output directly to the load Since a phase shift of 90° occurs in the artificial line a compensating phase shift of 90° is provided in the input circuit of one of the tubes to permit the outputs of the two tubes to add

Tube $V_2$ is so biased that only exciting voltages in excess of the normal carrier amplitude are amplified by it. $V_1$ operates as an ordinary linear amplifier whose output voltage is proportional to the excitation voltage. When the exciting voltage exceeds the carrier amplitude, both tubes supply energy to the load. This causes the equivalent impedance of the load end of the artificial line to be increased. The impedance of the sending end of a quarter-wave transmission line is equal to $Z^2/Z_L$ (see Chap. XV), where $Z$ is the characteristic impedance of the line, and $Z_L$ is the equivalent load impedance at the receiving end of the line. Consequently, when both tubes are delivering energy to the load, the impedance presented by the sending end of the line to the plate of tube $V_1$ is *decreased*. Tube $V_1$ is thereby enabled to deliver more output power

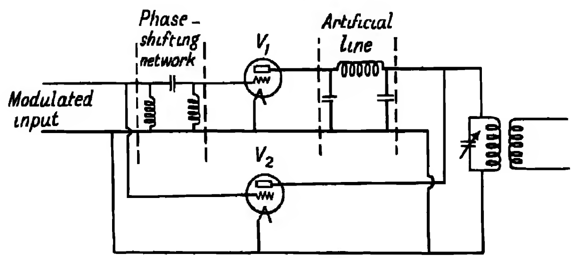



FIG. 210 The Doherty high efficiency linear amplifier

with the same a-c plate component. The amplification is substantially linear, and the over-all plate efficiency is high, normally exceeding 60 per cent.

## FREQUENCY MODULATION

Instead of varying the *amplitude* of the carrier frequency wave, it is possible to transmit intelligence by varying the *frequency* of the radiated carrier and maintaining the amplitude constant. The reader is referred to Chap. XII for a discussion of the nature of a frequency-modulated wave.

The simplest way to produce a frequency-modulated wave is to vary the capacitance of the transmitter-oscillator tuned circuit. An auxiliary condenser may be employed for this purpose. One plate of the capacitor is a thin diaphragm which is vibrated by a-f currents in much the same manner as a telephone receiver (see Chap. XIV). The resultant variation in capacitance causes the frequency of the oscillator tank circuit to vary in accordance with the a-f currents. In high-power oscillators, the capacitor-diaphragm mechanism is placed in an evacuated chamber to facilitate handling of high voltages.

The variable-impedance characteristic of an eighth-wave transmission line has also been used as a method of frequency-modulating a transmitter. If the transmission line is one-eighth wave length long, the

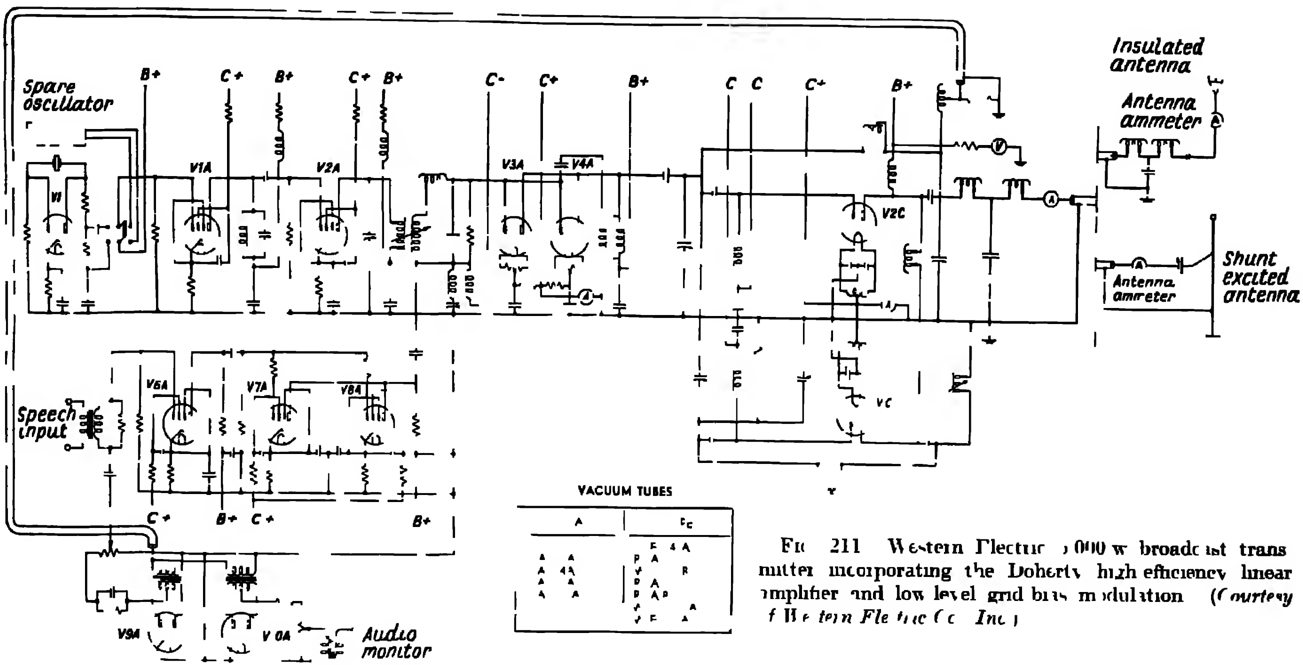



FIG. 211 Western Electric 5,000 w broadcast transmitter incorporating the Doherty high efficiency linear amplifier and low level grid bias modulation (Courtesy of Western Electric Co, Inc)

sending-end reactance varies with the resistance of the receiving end. The plate resistance of a tube is employed to provide the receiving-end resistance. When an a-f signal is impressed upon the tube grid, the varying plate resistance causes the sending end reactance also to vary in accordance with the a-f signal voltage. This varying reactance is made a part of the transmitter-oscillator tank circuit, which is thereby caused to vary in frequency.

A similar method of frequency modulation utilizes an inductance, capacitance, and resistance in parallel inductively coupled to the frequency-controlling circuit of a self-excited oscillator. Variations in the parallel resistance reflect reactance and resistance changes in the oscillator inductance and thereby cause the oscillator frequency to vary. The

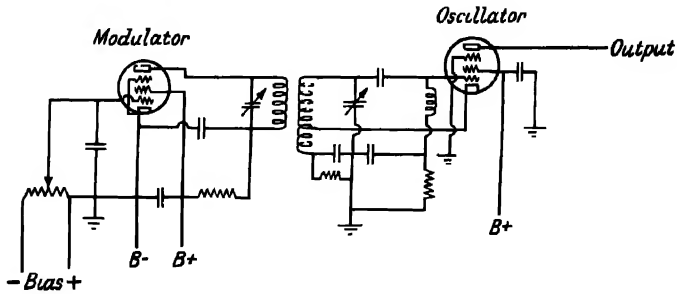



FIG. 212 Reactance-variation frequency modulation system

variable parallel resistance is supplied by the plate circuit of a vacuum tube, the input circuit of which is connected to the a-f modulating signal. Variations in tube plate resistance are accentuated by an additional resistance in the plate circuit. The simplified circuit is shown in Fig. 212.

In the Armstrong method of frequency modulation, the current from a relatively low frequency crystal controlled oscillator is phase-shifted. This phase shift represents a linear but very small frequency change that must be multiplied through a series of frequency doublers to produce the required frequency change in the radiated wave. In a typical system, an initial frequency of 200 kc is multiplied through a series of six doublers to 12.8 megacycles, where it is heterodyned down by a second crystal-controlled oscillator to a frequency one forty-eighth of the frequency to be radiated. For the 40-megacycle band, this frequency is of the order of 900 kc, which is then multiplied by four doublers and a tripler up to the 40-megacycle band.

Another method of frequency modulation makes use of the fact that the plate circuit of a tube will appear to be a reactance if some of the plate voltage is fed back to the grid through a phase-shifting circuit. The plate would present a capacitive reactance, for instance, if the phase-shifting circuit caused the grid voltage to lag the plate voltage by 90°.

The magnitude of the reactance varies with the gain of the tube. The frequency-modulated oscillator uses a variable-mu tube as part of the capacitance in the tuned circuit. An a-f voltage applied as a bias to the tube causes variations in the gain of the tube and, hence, in the frequency of the oscillations.

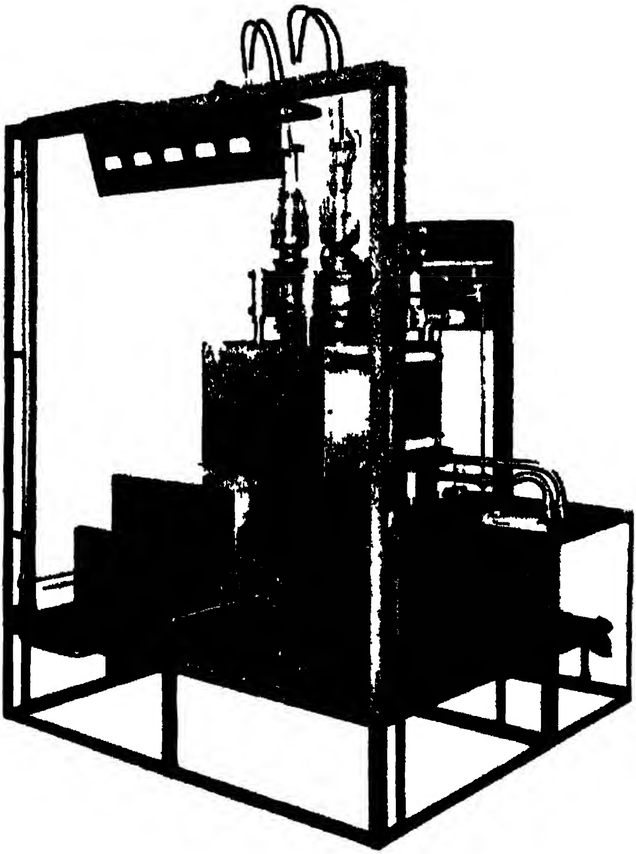


FIG. 213. The 50,000-w. final amplifier of a frequency-modulation broadcast transmitter operating on ultrahigh frequencies. Note the hose connections to the water-cooled tubes. In addition to the water cooling facilities the tubes are provided with air blast on their end seals. (*Courtesy of Radio Engineering Laboratories, Inc.*)

## *TRANSMITTER EMISSION CLASSIFICATIONS*

The emission of various types of radio transmitters has been classified according to international law. The classifications are based on the

purpose for which the wave is used and apply only to amplitude-keyed or amplitude-modulated signals. The wave forms for the various classifications are shown in Fig. 216. The classifications are as follows:

*Type* A0. In this classification are waves that have successive oscillations which are identical under fixed conditions. An unmodulated carrier wave would fall in this group.

*Type* A1. Telegraphy on pure continuous wave constitutes type A1

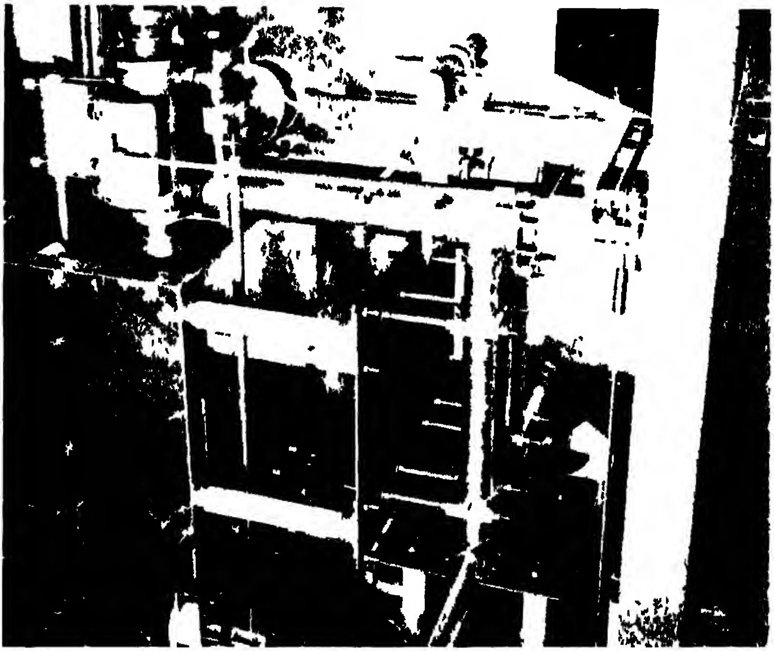


FIG. 214. Close-up view of 50-kw amplifier plate circuit. The antenna coupling inductance consists of a single hairpin turn of wire. (*Courtesy of Radio Engineering Laboratories, Inc.*)

emission. A type A0 carrier that is broken up or keyed, according to a telegraphic code is type A1 emission.

*Type* A2. Modulated telegraphy constitutes type A2. A keyed carrier wave modulated either by chopper or some other means at an audio frequency constitutes A2 emission.

*Type* A3. Radiotelephony (amplitude-modulated) constitutes A3 emission. This classification describes the wave form that results when a constant carrier is modulated by audio frequencies corresponding to voice, music, or other sounds.

*Type* A4. The signal produced by facsimile transmission comprises type A4. The reader is referred to Chap. XIX. Type A4 emission refers

to the wave form that results from the modulation of a carrier wave by frequencies produced at the time of the scanning of a fixed image with a view to its reproduction in a permanent form

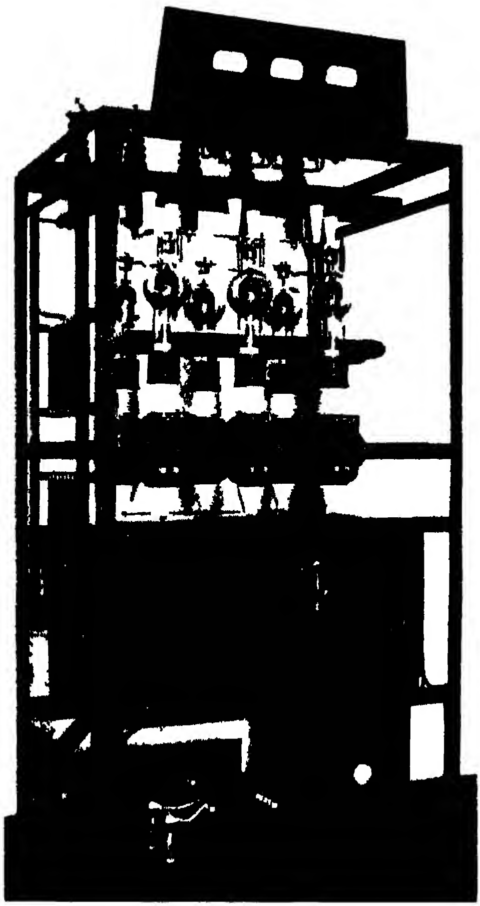


Fig 215 The 15,000 v rectifier unit which supplied plate power to the 50 kw amplifier of Fig 213 (*Courtesy of Radio Engineering Laboratories Inc*)

*Type* A5 The signal produced by a television transmitter (see Chap XIX) forms type A5 The wave form resulting from the modulation of a carrier wave by frequencies produced at the time of the scanning of fixed or moving objects is included in type A5 emission

*Type* B Type B emission comprises the wave form of a radiation composed of successive series of oscillations the amplitudes of which,

# Chapter XIV

# SOUND CONVERSION

The conversion of energy in the form of sound waves to electric energy, and vice versa, forms an important operation in numerous types of radio equipment, both transmitting and receiving. In transmitting systems, sound-wave energy is converted into electric energy by means of *microphones*. By means of the familiar *electrical transcriptions*, or phonograph records, sound waves are utilized to cut indentations, or tracks, in a plastic substance to form a permanent record of the sound-wave vibrations. By the use of a *pickup head*, these recordings can be made to supply electric energy corresponding to the original sound waves, which can be amplified and reconverted to sound waves or utilized to modulate a radio transmitter.

In receiving systems, electric energy in the form of a-f alternating currents is converted into sound by means of *telephone receivers* or *loudspeakers*.

Sound is essentially vibratory in origin. Anything that can be made to vibrate physically will set up sound waves in the surrounding medium. If the vibrations occur at an audio frequency, that is, at a rate within the range to which the human ear is sensitive, the resulting sound waves will be heard. Many forms of vibrations set up sound waves that are *beyond* the range of the human ear and, hence, are unheard. Nevertheless, these sound waves *do* exist and are often picked up by sound conversion devices, such as microphones, that are more sensitive to certain frequencies than is the human ear. Such sound waves, therefore, often create electric energy (through conversion devices) that results in apparently poor fidelity from this unsuspected source.

## *CONVERTING SOUND INTO ELECTRIC ENERGY*

**The Microphone Diaphragm.** All forms of microphone depend for their operation upon a **diaphragm** of metal that vibrates when sound waves impinge upon it. Since sound waves are vibratory in origin, they will, if strong enough, set up a physical vibration in any object that is free to vibrate.

If a tin pan is struck with a hammer, a loud sound results. The sound emanating from this source is caused by sound waves set up by the vibration of the tin pan as a result of the blow. The *frequency* of the

sound waves depends upon the physical dimensions of the pan, since this determines the *period* of the vibrations, that is, the length of time required to complete one vibration. The *volume* of the sound is a function of the *amplitude* of the vibrations, which in turn, depends upon the force with which the hammer strikes the pan.

If another tin pan of the same dimensions is in the vicinity of the original pan, *sympathetic* vibrations will be set up in this pan, and if one listens very closely, a sound will be heard emanating from this second pan which has the same tone as the sound from the first pan. Since their dimensions are identical, both pans have the same resonant, or natural, period of vibration. The second pan is therefore free to vibrate at the same frequency as the first pan, the energy causing the vibration being supplied by the sound waves striking the second pan. The amplitude of the vibrations will depend upon the strength of the sound waves striking the second pan and will therefore be greater as the second pan is moved closer to the first.

If the second pan is so mounted, or suspended, that it vibrates very easily and freely, it will be found that it can be caused to vibrate independent of the frequency of the original sound waves. Thus, a pan having dimensions that differ from the original, or struck, pan, will vibrate at the same frequency as the sound waves striking it, if so mounted. However, it will be found that the vibrations are greatest in amplitude when the two pans have the same natural period, that is, when their dimensions are identical.

From the above, it can be seen that the amplitude of sympathetic vibrations in the second pan is a direct function of the amplitude of vibrations of the first pan, except for *one* special case where two pans have identical physical dimensions and hence identical natural periods of vibration. In this case, the amplitude of the vibrations of the second pan are much greater.

If a thin metal disk is so dimensioned that its natural period of vibration does not correspond with any frequency within the a-f range, vibrations set up in it by sound waves will be essentially directly proportional to the amplitude of the sound waves, that is, the volume of the sounds. Such a metal disk when mounted so that it can vibrate very easily and freely is called a **diaphragm.**

**The Carbon Microphone.** The most elementary form of microphone is the carbon microphone. In it, a diaphragm is arranged so that it bears against a quantity of carbon granules in a small cylindrical container called the microphone **button.** A drawing of an elementary carbon microphone is shown in Fig. 217. The contact resistance between the various granules of carbon varies with the pressure exerted upon them, since this pressure determines the extent of contact. The button is an insulation material with electrodes at each end. In many types, the

diaphragm forms one electrode, or terminal. The conducting path between the electrodes depends upon the total resistance of the carbon granules. As the diaphragm vibrates because of the impingement of sound waves upon it, pressure is alternately impressed upon and removed from the granules. As a result, the resistance between the electrodes of the button varies at a rate corresponding to the frequency of the diaphragm vibrations The *extent* of this resistance variation depends upon the *difference* in pressures exerted upon the granules; that is, it is determined by the *amplitude* of the diaphragm vibrations. If the carbon button is connected in series with a battery, the direct current that flows will vary in amplitude in proportion to the volume of sound reaching the microphone and at a rate corresponding to the frequency (or tone) of the sound waves.

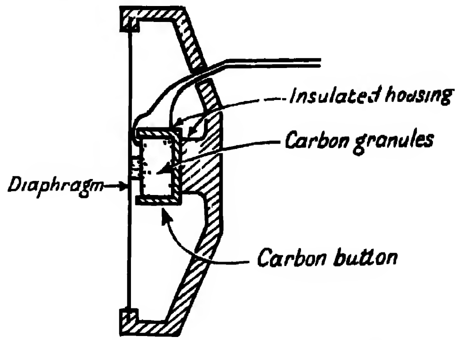



FIG. 217 Construction of elementary single-button carbon microphone

This varying direct current is applied to the primary of a transformer. The alternating voltage resulting across the secondary is applied to an a f amplifier of conventional design. The alternating voltage is thereby amplified until sufficient voltage is developed to drive a power amplifier and loudspeaker (in the case of a public address system), to supply modulation to a transmitter, and so on

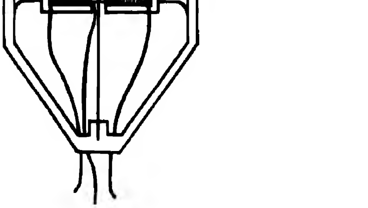



FIG. 218 Construction of elementary double button carbon microphone

An improved type of microphone, shown in elementary form in Fig. 218, utilizes *two* carbon-granule buttons and is called a **double-button carbon microphone.** A movement of the diaphragm in a given direction causes a compression of the granules in one button and a simultaneous expansion of the granules in the other button. Consequently, as the resistance of one button decreases, the resistance of the other button increases. This condition permits connecting the two buttons in the push-pull arrangement shown in the diagram with resultant greater sensitivity and output.

The microphones used in ordinary telephone communicating systems

are of the single-button carbon type. In such usage they are commonly called **transmitters,** but the term **microphone** is preferred in radio work in order to avoid confusion with other types of transmitters.

The carbon microphone is seldom used in broadcast stations, public-address systems, or other radiotelephone work where fidelity is important, owing, in part, to the comparatively poor over-all frequency response and to the steady background hiss that such units develop. The background hiss results from unavoidable random changes in resistance of the carbon granules. The later types of microphones, which are discussed below, although they have much lower sensitivity than the carbon microphone, have a much improved frequency response with correspondingly greater fidelity. Carbon microphones are used in police and amateur radio systems where sensitivity is of greater importance than fidelity, since only speech is transmitted.

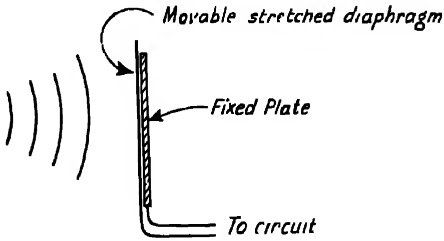


FIG. 219 Fundamental principle of capacitor microphone.

**The Condenser or Capacitor Microphone.** The constructional details of an elementary condenser or capacitor microphone are shown in Fig. 219. The diaphragm in this type of microphone acts as one plate of a capacitor, the other plate of which is fixed. The diaphragm is usually constructed of an aluminum alloy about 0.001 in in thickness, which is tightly stretched to avoid mechanical resonances (natural periods of vibration) at undesired frequencies. A d-c emf of several hundred volts is applied through a suitable resistor across the capacitor plates. The movement of the diaphragm as it vibrates at an audio frequency causes corresponding changes in the capacitance of the capacitor. These alter the charge of the capacitor and the consequent current flowing through the series resistor. The $IR$ drop across this resistor therefore varies in accordance with the sound waves impinging upon the diaphragm. This voltage is applied to the grid circuit of an amplifier.

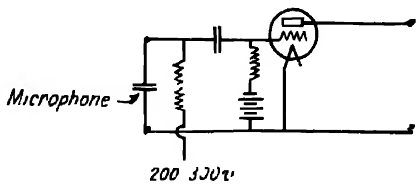


FIG 220 Circuit connections of capacitor microphone.

The spacing between the diaphragm and the fixed plate of the capacitor is made as small as possible, with the result that the capacitance change caused by movement of the diaphragm becomes an appreciable portion of the total capacitance of the condenser. Spacing of 0.001 in. is made

possible at the voltages employed by sealing the space between the plates from the outside air to keep out dust and dirt.

The capacitor microphone is a high-impedance device and may be coupled directly to the input circuit of the first amplifier stage as shown in Fig. 220. The capacitance shunted across the microphone should be made as small as possible by proper arrangement of the circuit leads and selection of the proper tube. The smaller the total input-circuit capacitance, the greater the portion of the total capacity that is represented by a given change in the microphone capacitance, resulting in greater efficiency.

The capacitor microphone has a very good frequency response characteristic but relatively low sensitivity. In order to maintain low input-circuit capacitance and to avoid stray capacitance pickup, it is necessary to place the first amplifier tube as close to the microphone as possible. The usual practice is to place the amplifier tube immediately adjacent to the microphone right on the microphone stand.

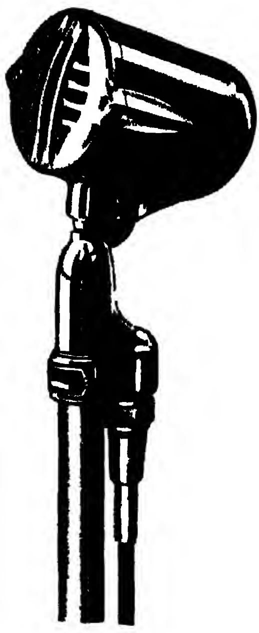


FIG. 221. Moving-coil (dynamic) microphone. (*Courtesy of Astatic Corp.*)

Because of its excellent frequency response, the capacitor microphone was at one time widely used in broadcast stations and public address systems. However, it has been largely replaced by later types for these applications, and its principal use at present is in making sound measurements.

**The Moving-Coil Microphone.** The moving-coil microphone has considerably greater sensitivity than a capacitor-type microphone of equivalent frequency-response characteristics. A moving-coil microphone (Fig. 221), also called the **dynamic microphone,** utilizes a diaphragm of nonrigid material to which is rigidly attached of a number of turns. Excellent l-f response is achieved because of the large displacement possible with the nonrigid diaphragm.

The coil is arranged to pass between the poles of a powerful cobalt-steel permanent magnet. As the coil moves with the diaphragm in response to sound waves, the magnetic lines of force between the magnetic poles are cut, causing a current to be induced in the coil. The clearance

between the coil and the magnetic poles is kept as small as possible in order to provide a magnetic field of maximum intensity. The coil current varies in frequency and amplitude in accordance with the diaphragm movement.

In order to provide good h-f response, the entire movable portion of the microphone must be as light in weight as possible. For this reason, the coil is usually constructed of thin insulated aluminum ribbon instead of copper wire.

In addition to the good sensitivity and excellent frequency response, the moving-coil microphone has the added advantage of low-impedance output. This characteristic permits the use of long cables between microphone and first amplifier without adverse pickup or frequency-response variation. In order to keep stray pickup at an absolute minimum, it is customary to use twisted-pair cable having a well-grounded shield.

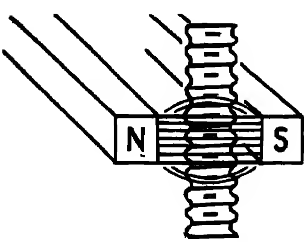



FIG. 222. Fundamental principle of ribbon microphone.

**The Ribbon Microphone.** The ribbon microphone is essentially a special form of moving-coil microphone inasmuch as electromagnetic induction is responsible for the emf developed. Figure 222 illustrates the constructional details. A small flat piece of aluminum-alloy foil is suspended between the pole pieces of a powerful permanent magnet. The spacing between ribbon and pole pieces is maintained as small as possible without interfering with the freedom of vibration. Unlike the arrangement in other types of microphones, *both* of the flat surfaces of the ribbon are exposed to the sound waves. The amplitude of ribbon vibration is therefore a function of the *difference* in sound pressure exerted on the front and back of the ribbon. When the distance between the two sides of the ribbon is appreciably less than a quarter wave length, the ribbon variation is directly proportional to the frequency and to the difference of pressure in sound waves between the two sides of the ribbon. Since the difference of sound-wave pressure is a function of the wave velocity, this type of microphone is often called a **velocity microphone.** By proper choice of ribbon dimensions, undesirable physical resonances are avoided.

By cutting away portions of the pole pieces, the distance between the front and back of the ribbon can be kept shorter than a quarter wave length at any of the frequencies that it is desired to use. As a result, uniform frequency response to beyond 10,000 c is possible with this type of microphone. Since the velocity of sound waves in air is approximately 1,100 ft per second, it is apparent that the distance between the front and back of the ribbon must be kept less than approximately $\frac{10}{32}$ in. to

provide the above response. This distance corresponds roughly to a quarter wave length at 10,000 c

The velocity microphone has the disadvantage of overemphasized l-f response when placed too close to the sound source. This, however, is practically the only disadvantage. It has the overwhelming advantage over other microphones of extremely good frequency response, in addition to simplicity and low output impedance with attendant coupling advantages.

One of the major advantages of the ribbon microphone for certain types of work is the pronounced directional characteristic. Since the microphone depends for its operation upon the difference in sound pressure between the two sides of the ribbon, it is obvious that this difference in pressure is greatest when the sound is directed at right angles to the plane of the ribbon. Sound directed to the microphone in the plane of the ribbon reaches both sides of the ribbon at the same instant, producing no pressure differential. The approximate directional characteristics are shown in Fig 223.

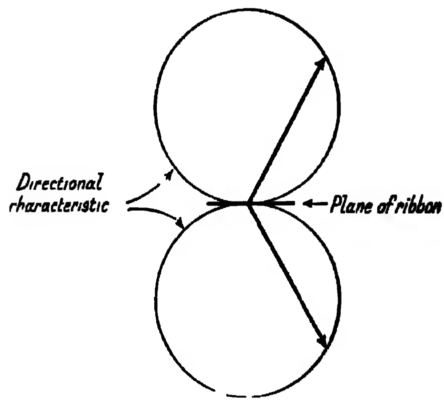


FIG. 223 Approximate directional characteristic of ribbon microphone

**The Crystal Microphone.** The piezoelectric effect of certain crystalline

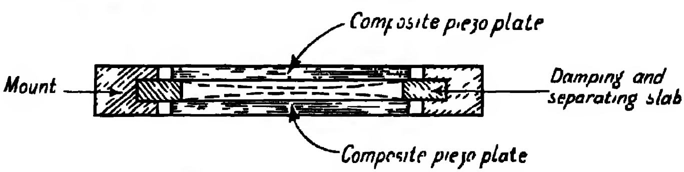


FIG. 224 Crystal microphone bimorph element

substances has already been discussed in Chap. XI. This effect in Rochelle salt crystals is utilized to transform mechanical stresses produced in the crystal by sound waves into electric energy. Perhaps the most successful type of crystal microphone consists of two Rochelle salt slabs arranged as shown in Fig. 224. This arrangement of crystals is known as a **bimorph element** and reduces the effect of temperature changes upon the crystals to a minimum.

When sound pressure is exerted on the conical diaphragm, the transmission of the pressure to the crystals causes a curvature of the crystals

in opposite directions because of the manner in which they are cut. This double curvature causes an emf to be developed across the crystals that is in direct proportion to the frequency of the sound wave. Substantially uniform frequency response is obtained with this type of microphone up to 15,000 c. The electrical output is very small but is usually increased in commercial microphones by connecting several crystal units in series within a single microphone standard. By making the entire crystal sound-cell assembly physically small, the natural period of the unit is made higher than any frequency that it is desired to reproduce, and physical resonances are avoided.

**The Electrical Transcription.** An important form of sound-conversion device is the electrical transcription, or recording. By means of such transcriptions it is possible to make permanent records of an aural sequence, such as a radio broadcast program. The program can then be reconverted from the record to sound any number of times.

Broadcast-station sponsors often have commercial radio programs recorded. A number of records are made from the original, or master, recording. These records are then distributed to various radio stations for broadcasting purposes. This system has the advantage over chain broadcasting of avoiding the expensive telephone line charges entailed in a chain network. Such recorded programs, however, cannot be broadcast simultaneously with the original program, and, hence, they are not used where a program has a time utility, such as news broadcasts and sports events.

Commercial recording is accomplished by amplifying the output of one or more microphones used to pick up the sound waves. The output of the amplifier is fed into a *cutting head* In commercial applications cutting heads are of two general types electromagnetic and crystal and operate to convert the electric energy from the amplifier into mechanical motion of a needle, called the **stylus.** The stylus point is made of a very hard material, usually diamond or sapphire, to eliminate wear. The mechanical vibrations of the stylus occur in accordance with the frequency of the electrical energy fed to the cutting head and are utilized to cut corresponding grooves in circular disks of plastic composition called **records.** The records used for original commercial recordings are composed of wax or commercial nitrocellulose.

The soft wax or plastic record is made electrically conductive by spreading a layer of extremely fine conducting powder, such as graphite, over its surface and is then electroplated. After a series of operations, a metal mold, or *stamper*, is produced from the original electroplated disk. From the stamper, the finished records are pressed out of plastic material. This material, when it has hardened, can be used for a great number of reproductions. From a single stamper, it is possible to obtain as many as 1,000 finished pressings.

The final records are utilized to reproduce electric energy corresponding with the original sound waves by means of a *pickup head*. Commercial pickups are of two types, electromagnetic and crystal, and perform the reverse function of the cutting head. When the record is rotated by means of a turntable, a stylus that is made to track in the record grooves is caused to vibrate by the groove undulations. These mechanical vibrations of the stylus are converted into electric energy of corresponding frequency by the pickup in much the same manner as the vibration of a diaphragm is converted into electric energy by a microphone.

**Recording.** Commercial recording equipment consists of two major components, the turntable with associated feed mechanism and the cutting head. The turntable actually consists of a miniature lathe designed to rotate the record at a uniform speed and to provide constant radial movement of the cutting head. Recordings are made at standard speeds of $33\frac{1}{3}$ or 78 rpm. The radial movement of the cutting head is usually accomplished by means of a threaded shaft coupled to the turntable driving motor. As a result, the stylus cuts a spiral groove in the record and prevents adjacent grooves from overlapping. With some systems, the cutting is started near the outer edge, or perimeter, and moves inward toward the center. Other systems start cutting near the center and work outward to the perimeter.

There are two methods of cutting records, namely, lateral cutting and vertical cutting. In lateral cutting the undulations are cut in the *sides* of the grooves. In vertical cutting, the undulations are cut vertically in the center of the groove. Only one type of record, the lateral-cut record, is in general use today, and this type will be discussed here.

The grooves in a disk record are usually spaced about 92 per inch. This allows a spacing of 0.011 in. from center to center of adjacent grooves, of which 0.006 in. is the width of the groove itself. This limits the maximum lateral displacement of the stylus to 0.0025 in., and in general a lateral motion of 0.002 in. is maintained.

Lateral recordings may be made in either of two ways: by the *constant-amplitude* system or by the *constant-velocity* system.

In constant amplitude recording, constant sound level for all frequencies at the microphone (assuming an over-all uniform frequency-response characteristic up to the cutting head) is represented by constant amplitude of the grooves cut in the record, that is, all recorded frequencies have the same amplitude.

In constant-velocity recording, constant sound level under the same conditions is not represented by constant amplitude of the record grooves, but by constant vibrational velocity. In this case

$$\text{amplitude} = \frac{\text{velocity}}{\text{frequency}} \qquad (1)$$

Since the velocity is constant, it follows that the amplitude is inversely proportional to the frequency.

Wave patterns which represent groove undulations for both types of recordings are shown in Fig. 225. For purposes of illustration, these patterns are shown with progressively increasing frequency. In the

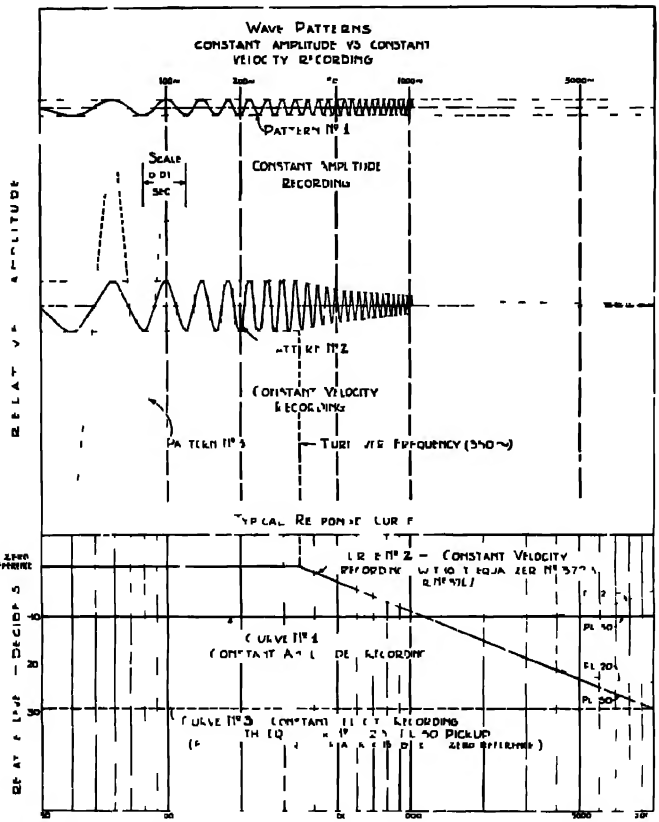


FIG. 225. Wave patterns for recording constant amplitude vs constant velocity. (*Courtesy of Brush Development Co.*)

constant-amplitude pattern it will be noted that for a constant sound level the amplitude is constant, regardless of frequency. In the constant-velocity pattern it can be seen that for constant sound level the amplitude increases as the frequency decreases. A frequency of 100 c in this system will have twice the amplitude of a frequency of 200 c.

Since this would necessitate excessive amplitudes at the lowest frequencies to obtain sufficient amplitudes at the higher frequencies during recording, and since these excessive amplitudes would severely limit the

number of grooves that could be recorded without crossover or echo effect, commercial constant-velocity records are cut with constant amplitude below 350 c and constant velocity above 350 c. This is indicated in Fig. 225. The critical frequency (350 c) is referred to as the **turnover point.**

In the electromagnetic type of cutting head, the a-f alternating current from the amplifier output is applied to the coil of an electromagnet. Between the poles of the electromagnet, as shown in Fig. 226, an iron armature is mounted on a pivot so that it is free to vibrate. The cutting stylus is rigidly attached to the armature. The vibration of the armature, and hence of the stylus, varies in frequency and amplitude in accordance with the alternating current flowing through the magnet windings. The damping block shown in the diagram acts to dampen, or decrease, excess vibration due to inertia of the armature and reduces the effect of mechanical resonances. Since the wax disk material offers very little opposition to the movement of the stylus, the damping block also acts to increase the load on the cutting head and brings the unit up to its proper terminal impedance.

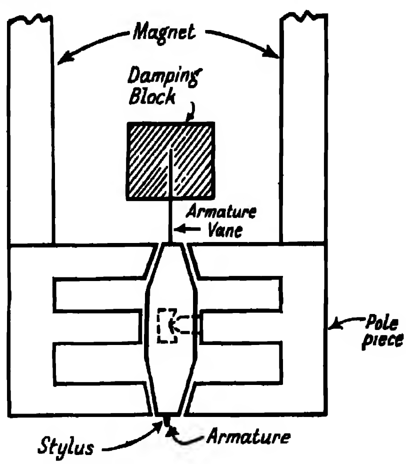


FIG. 226. Constructional details of magnetic type cutting head.

The crystal cutting head utilizes the piezoelectric effect of Rochelle-salt crystals to convert a-f alternating current to mechanical motion. When the emf is applied across the crystal, mechanical expansion and contraction of the crystal take place, which vary with the polarity and hence with the frequency of the applied emf. The crystal element is rigidly coupled to the cutting stylus, thus imparting a corresponding motion to the stylus. Some commercial units utilize four-ply crystal elements which actuate the cutting stylus through a stiff driving linkage and rocker-arm assembly. Because of the leverage obtained through the rocker-arm pivot assembly, a four-to-one step-up ratio in movement is obtained as measured from the crystal element to the stylus tip. Crystal cutting heads are characterized by uniform frequency response up to 10,000 c.

**The Reproducer or Pickup Head.** The process of the reproduction of a recording performed by the pickup head is the reverse of that of the cutting head; that is, it consists of converting the mechanical movements

of the pickup stylus into electric energy of corresponding character. The movement of the stylus is caused, of course, by allowing it to bear with slight pressure against the undulated grooves of the record which is rotated by means of a turntable.

The operation of the electromagnetic pickup is the reverse function of the principle of operation of the electromagnetic cutting head. The general constructional detail is also the same so far as elementary principles are concerned. The pickup stylus is rigidly attached to a movable pivoted iron armature, which is suspended between the pole pieces of a permanent magnet having an electromagnetic winding, or coil. The movement imparted to the armature by the stylus results in varying the lines of force of the magnetic field. An emf is thereby induced in the coil which is amplified in the same manner as the output of a microphone. The frequency and amplitude of the induced emf are functions of the armature and stylus movement.

The crystal pickup head utilizes a bimorph crystal element that is essentially similar to that of the crystal microphone. The pickup stylus is rigidly attached to the crystal element in such a manner that the movement imparted to it by the undulated record grooves varies the pressure exerted on the bimorph element. The output of this unit is consequently not subject to the effects of mechanical inertia, and for this reason the unit is often called an **astatic pickup.** The term indicates that the pickup as a body is in neutral equilibrium and hence has no tendency toward any change in position with resultant effects on the electrical output. The alternate compression and expansion to which the bimorph element is subjected by the movement of the stylus produce an emf across the element. The frequency and amplitude of this emf are essentially directly proportional to the frequency and amplitude of the stylus displacements.

The crystal bimorph element is sealed in a moistureproof cartridge. Pickup heads of this type are sensitive instruments and should not be subjected to rough usage. Any severe mechanical shock, such as would result from dropping the pickup arm or allowing the stylus point to hit the edge of the turntable, can permanently damage the unit. It is also important not to subject crystal pickup heads to temperatures in excess of 120 to 125° F. Such temperatures, particularly over a period of several hours, may seriously damage the crystal element.

Most types of recording and reproducing systems require the use of equalization networks to provide substantially linear output. Such equalizers are discussed in the section on studio apparatus in Chap. XVIII and will not be enlarged upon here. Although the use of an equalizer ensures higher fidelity, there are instances where it is desirable to alter the frequency response to suit particular conditions. For example, in reproducing commercial recordings, it is often desirable to omit the

equalizer in order to accentuate the bass frequencies. Omission of the equalizer is especially preferred in situations where records are reproduced at levels considerably below those of the natural sounds as heard in the studio. Here the bass frequencies as heard by the human ear appear to be attenuated considerably more than the middle-range frequencies.

In cases where such a bass boost would be too great, but where complete equalization for commercial constant-velocity records to obtain a flat response is not preferred, a suitable tone control may be added to the equalizer network. Often it is desirable to replace an equalizer with other high-impedance equalizers to obtain particular results. Although it may be desirable to accentuate the high frequencies, this procedure is not recommended, since it increases the background noises of the usual record. *Scratch* filters utilized to filter out record background noise are discussed in Chap. XVIII. This background noise, sometimes called **surface noise,** is relatively more pronounced in the higher frequencies. It is produced principally by tiny irregularities, abrasives, dirt, and so on, which affect the recording stylus.

The pickup-head position with respect to the turntable is very important, since this directly affects the lateral displacement of the stylus. The pickup should be mounted in such a manner that the stylus extends $\frac{5}{8}$ in. beyond the center of the turntable, as shown in Fig. 227. In the case of a crystal pickup, the underside of the pickup arm should be $\frac{3}{8}$ in. above the turntable, with the stylus resting in the groove.

**Wire Recorders.** Another system of recording and reproducing sound which was popularized during World War II is "wire recording." Wire recording consists of magnetic recording on steel wire and tape. Although only recently developed in this country, the magnetic recorder was invented in 1898 by a Dane, Valdemar Poulsen. Attention was focused on the principle of magnetic recording in the United States by the military requirement for a recorder which could be operated tilted in any position, which could sustain severe vibration while operating, and which would provide records which were relatively nondestructible and required no processing after recording.

The magnetic recorder utilizes the audio frequency current output of a conventional amplifier having a microphone input to create a concentrated magnetic field through which the steel wire or tape is passed. As the steel wire passes through this field it becomes magnetized in much the same way that a solenoid magnetizes an iron armature. In this case, however, the wire is steel and hence retains its magnetism indefinitely after it has passed beyond the influence of the magnetizing field. In laboratory tests wire recordings have been played back over 100,000 times with a loss in sound level of only 4.5 db.

The degree of magnetization of the steel wire varies as a function of the intensity of the magnetic field at the instant of exposure. The fidelity

of such a system therefore depends to some extent upon the speed of transit of the wire. Fidelity is also a function of the axis of magnetization. All wire recording is longitudinal recording; that is, the magnetizing force is impressed longitudinally along the wire, parallel to the direction of motion. Since there is no way to prevent radial displacement (twisting)

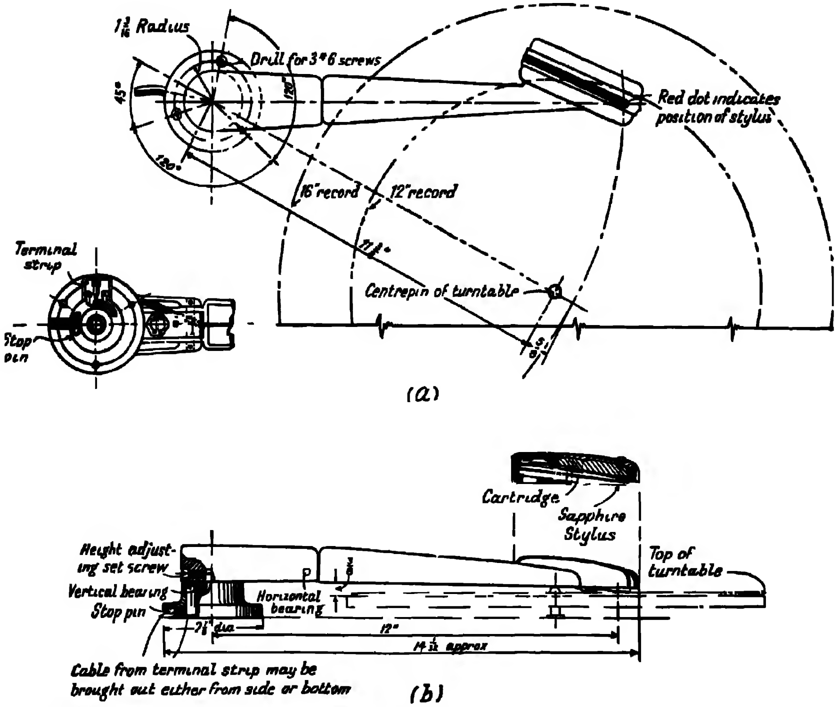



FIG. 227 Method of mounting cutting head (*Courtesy of Brush Development Co.*)

of a round wire during play back, longitudinal recording is the only method which can be used successfully with round wire.

With steel-tape recorders, transverse and perpendicular recording can be used as well as longitudinal. In transverse recording, the magnetic force is perpendicular to the direction of motion and parallel to the face of the tape. In perpendicular recording, the magnetic force is perpendicular both to the direction of motion and to the face of the tape.

Although tape is less subject to breakage than wire, wire is generally preferred because of smaller volume and weight requirements, and because a break can easily be spliced by tying a simple knot.

A complete wire-recording system consists of a microphone, a high-fidelity resistance-coupled amplifier, a loudspeaker, a recording head (which also serves as reproducer head), a demagnetizing or "erasing"

coil, and the associated driving mechanism to wind the wire from one spool to another. During recording, the microphone is coupled to the input of the amplifier. The output of the amplifier is fed directly to the recording head. The resulting magnetization of the wire varies in accordance with the intensity and frequency of the sound waves that impinged on the microphone diaphragm. Throughout the recording process, the erasing coil is energized so that the wire is certain to be free of any previous magnetization.

During playback, the reproducing head is coupled to the input of the amplifier. The output of the amplifier is fed to the loudspeaker. The erasing coil is de-energized. The moving magnetic fields created by the motion of the magnetized wire induce corresponding audio-frequency currents in the reproducer windings. These amplified currents energize the loudspeaker, which faithfully reproduces the sound waves in the usual manner.

Erasing is accomplished by energizing the erasing coil with a relatively high-frequency voltage, usually about 30,000 cycles, in the same manner as a jeweler demagnetizes a watch. This completely demagnetizes the wire as it passes through the erasing coil. The high-frequency voltage is supplied by an oscillator tube operating from the amplifier power supply.

The wire used in a magnetic recorder need not be at all large in diameter to serve its purpose. In the model 50A recorder manufactured by the General Electric Company, the wire is only .004 inch in diameter and is a specially heat-treated steel piano wire 11,500 feet, or approximately 2 miles, long. This entire spool weighs only half a pound and is good for over one hour of continuous recording.

During recording or playback, the take-up spool of a wire recorder is driven at constant speed, resulting in a gradual increase in the linear velocity of the wire past the recording head, due to the build-up of the wire on the spool. This speed-up is somewhat similar to the variation of the record-track velocity relative to the stylus in a conventional disk recorder as the stylus progresses from the perimeter of the disk to the innermost groove. However, in a wire recorder, this variation in velocity is of much smaller magnitude than in a disk recorder. In the General Electric recorder previously mentioned, the linear velocity of the wire varies only $+7$ percent from the nominal velocity of 2.9 ft/sec during an hour's recording. By way of comparison, the track-to-stylus velocity of the average disk recorder varies as much as $\pm 50$ percent from its nominal velocity during only 5 minutes of recording (12 inch record).

Since the magnetization curve for iron exhibits considerable curvature near its origin, direct magnetization of the wire by audio-frequency currents would result in considerable distortion. In some recorders this difficulty is overcome by first saturating the steel wire with a d-c field before applying the audio-frequency field. This has the effect of utilizing

the flat, or linear, portion of the hysteresis loop, in much the same manner that the proper bias voltage permits utilization of the linear portion of a vacuum-tube amplifier characteristic. In this type of system, both the d-c field and the a-f field are applied in the recording head by separate coils.

Other recording systems utilize a high-frequency voltage simultaneously applied to the common recording-head coil. The same 30,000-cycle voltage provided for the erasing coil is utilized for this purpose. This voltage has the effect of flattening out the hysteresis loop, apparently owing to the high-frequency excitation of the iron molecules making them more susceptible to realignment by the low a-f field.

The wire recorder has found many applications in the commercial field, both for home-radio recorder use and in the broadcast industry. It is especially useful in broadcast work for remote pickups because of its ruggedness and portability. A solid hour of program may be recorded on one record (wire) and is immediately available for playback and rebroadcasting. If desired, the recording can be erased and in a matter of a few minutes the wire is available for further use.

In the later models, fidelity exceeding that of many disk-type recorders can be obtained. Some recorders have two-speed drives. Low speed is used for speech recording. High speed (usually double the low speed) is used for music recording. Despite the fact that the permissible recording time is cut in half at high speed, linear frequency response up to 7,000 cycles per second is obtained in some instances.

## CONVERTING ELECTRIC ENERGY INTO SOUND

There are many ways in which electric energy can be converted into sound, and all of them make use of the alternating electric current to actuate a material body and make it vibrate. If the alternating current is of an audio frequency, the vibrations occur at an audio frequency and set up sound waves to which the human ear is sensitive.

There are two sound-conversion units of this type that are of interest to the radio industry, namely, the telephone receiver and the loudspeaker. Both units utilize some form of the principle of electromagnetic induction for their operation. Piezoelectric-operated units, however, do not utilize electromagnetic induction.

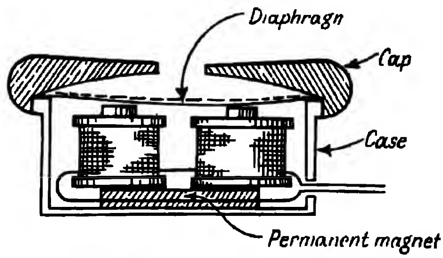
There are two problems connected with sound units of this type: the problem of handling sufficient audio power and producing sound waves of the desired amplitude (volume) and that of delivering the best possible fidelity signal.

**The Telephone Receiver.** Telephone receivers, so far as most present-day applications are concerned, are not essentially high-fidelity devices. There are, to be sure, many telephone receivers available today having excellent frequency-response characteristics. Most of these units,

however, are utilized in applications where high fidelity is not one of the essential requirements.

The primary requirements of receivers for modern use are that they be light in weight; that they deliver sufficient volume or signal when held close to the ear by means of appropriate headgear; and that their acoustical design permit effective elimination of external room noises. These receivers find wide application in radiotelegraph stations where a number of operators are receiving and copying different signals simultaneously. The confusion that would result from using separate loudspeakers for this purpose is apparent. "Headphones," "earphones," "receivers," and simply "phones" are all terms often used to describe telephone receivers and are synonymous, inasmuch as the same basic unit is concerned.

The most common form of telephone receiver is that operating on the electromagnetic principle, often called the **magnetic earphone.** This unit utilizes a small but powerful permanent magnet attached to pole pieces about which two coils are wound. The coils are wound with several hundred turns of very fine wire, and a circular metal disk of thin iron is so mounted that it forms a part of the magnetic circuit. The disk, called a **diaphragm,** is physically supported so that it is free to vibrate easily and is similar to the diaphragm used in some types of microphone.




228 Details of magnetic-type earphone.

When no current is flowing through the coils (which are connected in series), the soft iron diaphragm tends to cling to the pole pieces because of the attraction of the permanent magnet. When an a-f alternating current is passed through the coil, the resulting electromagnetic field that is set up alternately aids and opposes the permanent-magnet field as its polarity periodically reverses. Consequently, the attraction exerted upon the diaphragm alternately waxes and wanes, with resultant corresponding changes in the displacement of the diaphragm. The resultant vibration of the diaphragm is synchronized with the frequency of the current flowing through the coils, since the variation in field strength is a direct function of the frequency. The diaphragm vibration sets up sound waves of corresponding frequency in the air as previously described. Since the field strength varies directly as the amplitude of the coil current, the diaphragm displacement and subsequent sound-wave amplitude (or volume) are direct functions of the current amplitude. The constructional details of a magnetic-type earphone are shown in Fig. 228.

The sensitivity of a magnetic earphone is determined by the number

of turns of the coils and by the weight and flexibility of the diaphragm. For this reason earphone coils are wound with as many turns as possible. Despite the fact that a greater number of turns of wire in a given space (since space is limited in a receiver) necessitates a higher resistance winding (more wire of smaller size) and decreases the current, the receiver sensitivity is increased, since the ampere-turns are increased, as can be seen from the following illustration:

Assume that a receiver has coils wound with 1,000 turns of a given size of wire and a resistance of 500 ohms. When the receiver is coupled to a 10,000-ohm circuit, the total series resistance of the output circuit is 10,500 ohms. If the voltage across this circuit is 50 v, by Ohm's law the current will be

$$I = \frac{E}{R} = \frac{50}{10{,}500} = 0.0047 \text{ amp}, \tag{2}$$

and the ampere-turns will be

$$\text{ampere-turns} \quad 0.0047 \cdot 1{,}000 = 4.7. \tag{3}$$

If wire having one fourth the cross-sectional area is used, it will be possible to wind *four* times as much wire in the same space, or 4,000 turns. The resistance of the coil, however, will be increased approximately 16 times, so that it will now be 8,000 ohms. The total circuit resistance will therefore be 18,000 ohms, and the current that will flow in the circuit under these conditions is

$$I = \frac{E}{R} = \frac{50}{18{,}000} = 0.00278 \text{ amp}, \tag{4}$$

and the ampere-turns will be

$$\text{ampere-turns} \quad 0.00278 \cdot 4{,}000 = 11.1. \tag{5}$$

Since the sensitivity of the receiver is a direct function of the ampere-turns, the sensitivity of this receiver has been increased more than two times. High-resistance phones, in general, enable a better load matching to be obtained in radio circuits and are therefore to be preferred.

Another type of magnetic earphone, which at one time found fairly wide usage, is the *balanced-armature* type. This unit operates on a variation of the electromagnet principle described above. Although it still has a number of ardent exponents, because of its weight and bulkiness, this type of earphone has never attained popular favor and has been largely superseded by later types.

The *crystal earphone* is another application of the piezoelectric effect of Rochelle-salt crystals. In this application the expansion and contraction in the crystal caused by applied emf are utilized to displace a diaphragm similar to that of the magnetic receiver. Although crystal receivers are characterized by inherently better frequency response than

the magnetic types, they do not find such wide usage as the latter. The predominant practical characteristic of the crystal receiver is its light weight. However, magnetic receivers weighing only a few ounces have been developed, so this advantage of the crystal unit is largely counterbalanced. Crystal units have the additional disadvantage that they cannot be placed in series with d-c circuits, which necessitates the use of an output coupling transformer that in many cases is undesirable.

**The Loudspeaker.** Loudspeakers, as the name implies, were originally designed to obviate the necessity for wearing headphones. The aim was to reproduce the sounds of a program at comfortable room volume so that a number of persons could simultaneously enjoy the program. Such loudspeakers have since enjoyed a wide diversity of application in many additional fields. Modern loudspeaker design practice embraces a number of other qualities, including volume capability, such as required for public-address systems, and high fidelity. The ultimate ideal in loudspeakers used in radio work is to reproduce the sound wave that would have reached the ear of the listener if he had been present in the place where the original music or speech was produced. Such an ideal requires wide, uniform frequency-response characteristics and the ability to handle large as well as small amplitudes of power with equal fidelity.

Loudspeakers can be designed to radiate sound in a beam, similar to that sent out by an automobile headlight, or to radiate sound uniformly in all directions. The loudspeakers normally used in radio work usually have a radiation characteristic in the form of a hemisphere in front of the loudspeaker; whereas those used in theater talking-movie systems are made directional toward the audience, which reduces the effects of reverberation from the walls. Maximum radiation reaches the audience directly before reflections from other surfaces.

One of the earliest loudspeakers was of the **horn** type, which consisted of a magnified edition of the magnetic earphone connected to a long curved horn of proper exponential design. The radiation of such loudspeakers at low audio frequencies is poor. Radiation can be increased by making the diaphragm larger; but this necessitates making the diaphragm thicker to obtain sufficient rigidity, and the resultant added mass reduces the amplitude at high frequencies.

Horn-type loudspeaker installations are utilized for applications where space is not at a premium, such as outdoor public-address systems. For such applications, best results have been obtained by using a number of units having comparatively small diaphragms instead of a single unit with one large diaphragm. Although the low-frequency radiation characteristic of such a system is good, provided the horns are long enough, it has been found that using additional small units does not increase the h-f output above certain critical radiation levels.

The use of conical horns and of some types of flaring horns has the effect

of loading up the diaphragm to a point where the l-f radiation is materially increased. The flare of the horn is the factor that determines to how low a frequency the loading caused by the horn becomes effective. In order to obtain efficient radiation, it is necessary to make the mouth opening of the horn approximately equal to the wave length of the sound being radiated. To obtain the large-mouthed opening with the very small flare required for proper reproduction of low frequencies, the horn length becomes excessive and impractical for ordinary use.

A once extremely popular loudspeaker (still used for many applications) is the **cone**-type loudspeaker. The constructional details are shown in

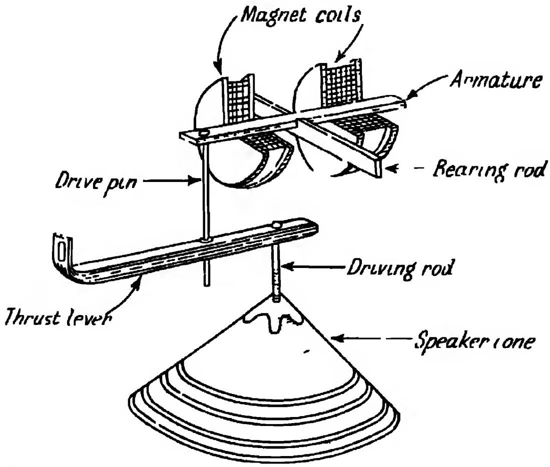



FIG. 229 Constructional details of cone type loudspeaker.

Fig. 229. Two magnet windings are wound about a spool mounted longitudinally between the pole pieces of a permanent magnet. A soft-iron armature is mounted on a pivot as shown in the illustration. Audio-frequency alternating current flowing through the coil windings varies the permanent-magnet field flux, alternately opposing and adding to the total flux. This varying field acting on the armature causes it to vibrate at a frequency corresponding with that of the alternating current in the coils. This vibration is transmitted to the loudspeaker cone through the driving pin, thrust lever, and driving-rod arrangement, and the vibrating loudspeaker cone sets up sound waves.

The most popular type of loudspeaker, which finds almost universal application today, is the **dynamic** type. Like the dynamic microphone, this unit operates on the same principle and utilizes a moving coil as its basic component. The constructional details of an elementary type of dynamic loudspeaker are shown in Fig. 230(a). The dynamic loudspeaker consists of a small coil rigidly attached to the peak of a paper loudspeaker

cone. The coil is so arranged that it is free to move longitudinally between the field poles of a magnet. To increase the total flux in the magnetic circuit a cylindrical iron core is mounted in such a manner that it passes through the center of the coil. The spacing between coil and pole pieces and the iron core and coil is kept as close as possible to ensure maximum flux density through the coil. The coil is kept from touching either the cylindrical core or the pole pieces by anchoring the center of the cone,

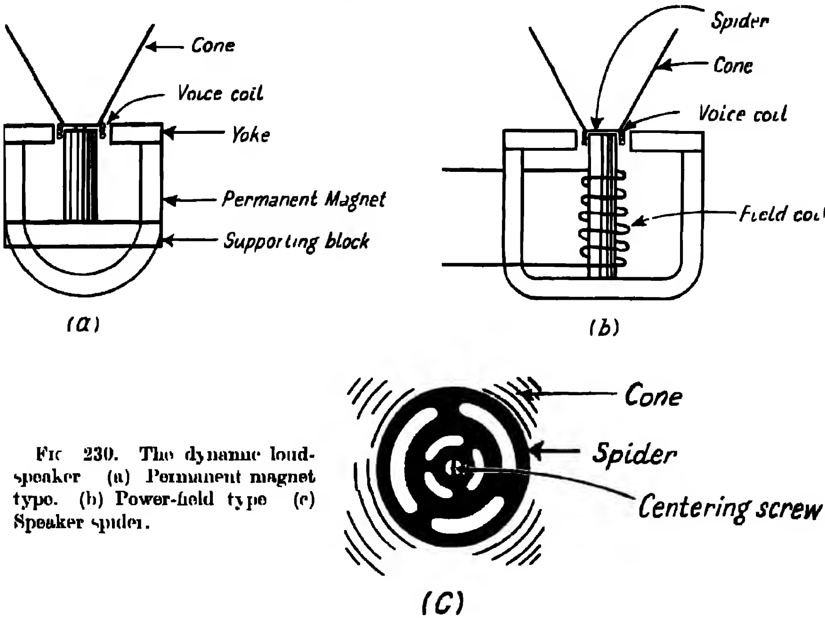



FIG 230. The dynamic loudspeaker (a) Permanent magnet type. (b) Power-field type (c) Speaker spider.

and hence the coil, to the center of the core by a thin piece of metal or heavy paper called the **spider,** as shown in Fig. 230(c). The spider is very flexible in a longitudinal direction, but it is stiff in the transverse direction. Thus it prevents lateral displacement of the coil, which might cause it to touch either core or pole pieces, without interfering with the desired longitudinal movement. Contact of the coil with other parts of the assembly would result in undesirable extraneous noises from the loudspeaker and would cause scratchy reproduction

The dynamic loudspeaker reverses the usual procedure of loudspeakers operating on the electromagnetic principle. Instead of passing the a-f alternating current through the field coils of the loudspeaker, this current is passed through the small movable coil described earlier. A powerful *fixed* magnetic field is maintained about the dynamic, or moving, coil. When a current is passed through the coil, the resulting coil field is either attracted to, or repelled from, the powerful fixed field, depending upon

the polarity. This attractive or repellent force causes a displacement of the coil along its longitudinal axis, with a corresponding movement of the loudspeaker cone. When an alternating current is passed through the coil, the displacement is in the form of a vibration caused by the alternate attractive and repellent forces exerted, and the rate of vibration is a direct function of the frequency. The velocity of the coil, and, hence, of the loudspeaker cone, is a direct function of the amplitude of the coil current.

Compared with other types, the dynamic loudspeaker has been found to be remarkably sensitive. A very small coil current will cause cone vibration due to the powerful attraction and repulsion of the fixed field. As a result, dynamic loudspeakers are capable of developing large volume compared with other loudspeakers of the same size. By proper physical dimensioning of the circular cone as to thickness and diameter, and through highly developed cone-mounting arrangements utilizing cone material of proper texture and flexibility, extremely high fidelity has been obtained.

Earlier types of dynamic loudspeakers utilized permanent magnets to supply the fixed field, or *power field*, as it is called. Modern dynamic loudspeakers obtain the power field by means of a large solenoid, or electromagnet, called the **field coil**, as shown in Fig. 230(b). A much more powerful field can be developed by this means, but such units have the disadvantage of requiring comparatively large amounts of direct current to supply the field coil. Nevertheless, because of the greater output, power-field dynamic speakers are in overwhelming use. Most receivers, amplifiers, and the like, utilize the power pack of the apparatus to supply the direct current for the loudspeaker field. In modern apparatus, this field is connected in series with the d-c plate supply and forms an integral part of the filter system. The field coil thus serves the double purpose of acting as a filter choke and supplying loudspeaker field power.

Permanent-magnet dynamic loudspeakers (p-m dynamics) are utilized for small receivers where space limitations prevent the use of the relatively bulky field coil or in applications where d-c power is not readily available. The latter application includes public-address systems where it is often necessary to use a number of loudspeakers at locations far removed from the source of power. Some very efficient p-m dynamic loudspeakers have been developed, which make use of recent developments in magnetic material.

Because of the small dynamic coil, dynamic loudspeakers are low-impedance devices and require the use of output coupling transformers to effect the proper impedance match. Typical dynamic-coil (often called **voice-coil**) impedances range from 2 to 12 ohms. Because of this low impedance, it would be impractical to connect such a loudspeaker directly to a power-amplifier circuit. With the high plate impedance of a typical power-amplifier plate circuit, the power transfer to the loudspeaker would be negligible.

Often objectionable hum is developed by power-field dynamic loudspeakers. When the trouble is not caused by the receiver or amplifier to which the loudspeaker is coupled, it is due to poorly filtered field current. If the loudspeaker field is part of the receiver filtering system, the field current is usually never exactly smooth but contains ripple components as discussed in the section on power packs in Chap. XII. Such variations in field current cause an emf to be induced in the loud speaker voice coil. This, in turn, causes displacement of the coil and cone and results in a 60- or 120-c hum being produced by the loudspeaker.

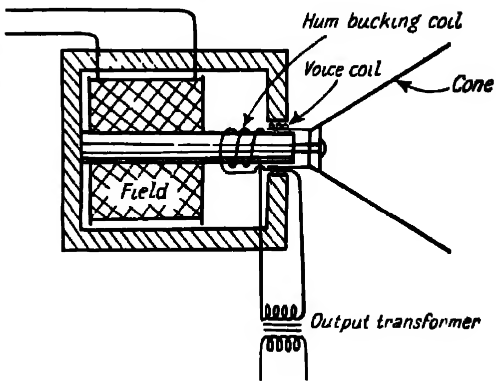



FIG. 231. Hum-bucking arrangement in dynamic loudspeaker.

Two systems have been developed to eliminate hum from this source in dynamic loudspeakers. One system utilizes a *hum-bucking coil*, and the other, a so called *shading ring*. The hum-bucking arrangement utilizes a small coil placed adjacent to the loudspeaker voice coil, as shown in Fig. 231. The coil is connected in series with the voice coil but is wound in the opposite direction. The design of the coil and its position in the loudspeaker are carefully arranged to make the hum voltage induced in it exactly equal to that induced in the voice coil. Although the hum-bucking coil has fewer turns than the voice coil, its position with respect to the field coil can be arranged so that equal hum voltage is induced in it. Owing to the series-opposing connection, the two hum voltages are canceled. Hence, no hum current flows in the voice-coil circuit, and no hum is produced by the loudspeaker. Because the hum-bucking coil is fixed to the core, however, the effect on the a-f output is negligible.

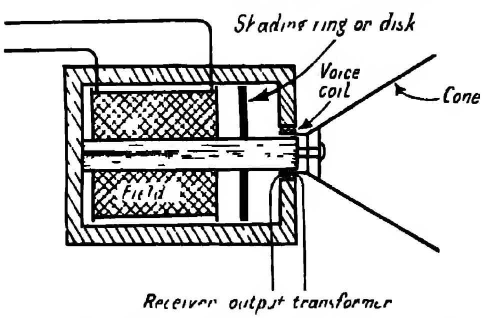



FIG. 232. Shading-disk arrangement for hum reduction.

The other hum-prevention arrangement consists of a thick copper disk placed between the voice coil and the field coil, as shown in Fig. 232.

This shading disk, or ring, acts as a single-turn coil of extremely low resistance in which strong eddy currents are induced by any fluctuation of the field magnetism. These eddy currents react on the main field, tending to oppose and to suppress such fluctuations and thereby effectively neutralizing them. Since this action prevents any field fluctuation from reaching the voice coil and inducing an emf in it, the copper disk is said to shade the voice coil. Because of its simplicity, this arrangement is widely used and has been found very effective

## *THE HUMAN EAR*

The final element in any sound system is the human ear, which hears the sound. In order to design a sound system that reproduces sounds with the ideal perfect fidelity, it is necessary to qualify the definition of the term "fidelity" in its relation to the human ear. Theoretically, it would appear that a system that affords an over-all linear frequency response would provide ideal fidelity. It should be remembered that over-all includes *all* units used between the actual source of the sound and the listener's ear. In a public-address system, microphone, amplifier, and loudspeakers would be included, and in a broadcasting system microphones, speech amplifiers, modulators, transmitter, receiver, and loudspeaker would all be included.

Actually, if the over-all response of such a system were linear through out, the sounds would probably not appear the same to a listener as they would if he were listening to the sounds at their actual source, because of certain peculiar characteristics of the human ear. Since the ultimate objective of any sound system is to reproduce a sound exactly as it would sound at its place of origin, a knowledge of the characteristics of the human ear is desirable and necessary.

**Characteristics of the Human Ear.** The normal ear recognizes tonal qualities in sounds varying in frequency from 16 to 16,000 c. These limits are, of course, approximate and vary from ear to ear. The frequency response of the human ear is definitely not linear. The ear sensitivity depends upon the frequency and has been found to be maximum for the average ear in the range of 1,000 to 3,000 c. Frequencies below 20 c are perceived by feeling rather than hearing, and frequencies above 16,000 c are not heard at all by most ears.

A sound wave in air is a compressional wave that is characterized by an increase and decrease of sound pressure above and below atmospheric pressure in the path of the wave. Actually, the pressure of a sound wave is the difference between normal atmospheric pressure and the atmospheric pressure that exists when the sound wave is present. The pressure of a sound wave is a function of the power developed by the loudspeaker or other sound-conversion device producing the

sound When a sound wave has a high pressure, it is said to have a **high intensity.**

The minimum value of sound pressure that gives a sensation of tone is called the **threshold of audibility.** The threshold of audibility varies with the frequency, being lowest through the middle range of frequencies (1,000 to 3,000 c) and highest at the very low and very high frequencies; that is, very low and very high frequency sounds must have fairly high intensities in order to be heard, whereas those in the middle-frequency ranges can be heard at much lower intensities.

The maximum value of sound pressure at which the sensation of tone is retained is called the **threshold of feeling.** Above the threshold of feeling, sound intensity becomes great enough to produce a sensation of pain rather than one of hearing. The threshold of feeling also varies with the frequency of the sound, being highest through the middle range of frequencies and becoming gradually lower as the very low and very high frequencies are approached. In other words, the human ear is capable of hearing sounds of higher intensity in the middle frequencies than at either the very low or very high frequencies. At the very low and very high frequencies, the sensation of hearing is lost, and the sensation of feeling starts at comparatively low intensities. A great deal of sound of high intensity can be heard at the middle frequencies before the threshold of feeling is reached.

Many peculiar psychological effects result from the foregoing characteristics of the ear. Thus, the frequency response of the ear varies greatly with the intensity of the sounds it is hearing. For example, the nonlinear response to sounds of great intensities causes harmonics and heterodynes of the original tones, which are not present in the original sound, to be produced. These tones, conceived within the ear and perceived by the brain, are called **subjective tones.** Often, when l f sounds of high intensity are heard by the ear, the subjective tones, or harmonics, produced by the ear interfere with the perception of higher frequencies. This phenomenon is called **masking** and is the reason why it is necessary to raise the voice when carrying on a conversation in a noisy location.

The magnitude of the auditory sensation produced by a sound is called its **loudness.** Because of the nonlinear frequency response of the ear and because of the variation of the threshold of audibility and the threshold of feeling with frequency, the loudness of a sound does not vary directly with its *intensity* The sensitivity of the ear has been found to vary approximately logarithmically with the intensity of the sound, and for this reason a logarithmic scale is used for measuring sound intensities.

The smallest *change* in sound intensity that is discernible to the human ear has been termed the **decibel.** Expressed as a function of the change in power required to produce this change in sound intensity, the decibel

may be defined as 10 times the common logarithm of a power ratio. Expressed mathematically,

$$N_{db} = 10 \log \frac{P_1}{P_2}, \tag{6}$$

where $N_{db}$ = number of decibels by which the sound intensities produced by $P_1$ and $P_2$ differ; and $P_1$, $P_2$ two different values of power expressed in common units.

The decibel, commonly abbreviated db, is one tenth of a bel and is often called a **voice unit.** The bel was universally adopted in 1928 and is named in honor of Dr. Alexander Graham Bell, inventor of the telephone. This unit is of great value in computing the gain in amplifiers in sound systems and in all types of a-f equipment, since it directly defines amplification in terms that are related to the sensitivity of the human ear. The input and output power, or over-all useful gain of an amplifier, can be calculated and expressed directly in decibels by means of Eq. (6). Conversely, when a given input power is available, such as from a microphone, and the desired decibel level is known, the necessary power output can be computed.

When the powers involved are expended across the same or equal resistances, the decibel difference can be expressed directly as a voltage ratio. This is derived as follows:

$$N_{db} = 10 \log \frac{P_1}{P_2}. \tag{6}$$

By the power equation,

$$P_1 = \frac{E_1^2}{R}, \tag{7}$$

$$P_2 = \frac{E_2^2}{R}. \tag{8}$$

Then Eq. (6) becomes

$$N_{db} = 10 \log \frac{\dfrac{E_1^2}{R}}{\dfrac{E_2^2}{R}}. \tag{9}$$

Where $R$ is the same in both cases, Eq. (9) becomes

$$N_{db} = 10 \log \frac{E_1^2}{E_2^2}, \tag{10}$$

$$N_{db} = 20 \log \frac{E_1}{E_2}. \tag{11}$$

Equation (11) is often useful in converting the gain of voltage amplifiers to decibels.

It should be remembered that the decibel is in itself not an absolute quantity but refers merely to a *change* in sound intensity, which by the above formulas can be reduced to a *change*, or *ratio*, between two powers, or voltages. In order to permit usage of the decibel as an absolute unit, *zero decibels* has been standardized to mean a power of 0.006 w (6 mw) expended in a resistance of 500 ohms. This standard facilitates the rating of components directly in decibels. Thus, if an amplifier is said to have an output of + 36 db, by Eq. (6) it can be calculated that the output is 36 db above 6 mw, or 24 w. Similarly, a microphone rated at - 50 db will deliver an output power of 50 db less than 6 mw, or $6 \cdot 10^{-8}$ w.

A chart listing decibel equivalents for voltage and power ratios will be found in Table IV in the Appendix

## *QUESTIONS AND PROBLEMS**

**1.** Low-impedance head telephones of the order of 75 ohms are to be connected to the output of a vacuum-tube amplifier. How may this be done to permit most satisfactory operation?

**2.** Which type of commonly used microphone has the greatest sensitivity?

**3.** Describe the construction and characteristics of a crystal-type microphone.

**4.** Describe the lateral system of transcription recording. Discuss the advantages and disadvantages.

**5.** What types of microphones have a high-impedance output?

**6.** What are the advantages of the single-button carbon microphone?

**7.** Draw a single schematic circuit showing a method of coupling a high-impedance loudspeaker to an a-f amplifier tube without flow of tube plate current through the speaker windings and without the use of a transformer.

**8.** Define the term decibel.

**9.** If a certain a-f amplifier has an over-all gain of 40 db and the output is 6 w, what is the input?

**10.** If the power output of a modulator is decreased from 1,000 to 10 w, how is the power loss expressed in decibels?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

# Chapter XV

# ANTENNAS

The phenomenon of electric waves being radiated into space is associated with the mutual interaction between electric and magnetic fields. A full understanding of this phenomenon can be predicated only on a thorough knowledge of the behavior of alternating currents. The student is urged to review the chapters on basic a-c theory before proceeding with the study of antenna theory

## *THE THEORY OF WAVE PROPAGATION*

An analysis of the fundamental theory underlying wave propagation goes back to the basic laws for electromagnetic induction (sometimes called the basic laws for generator action). These laws are:

1. *A moving electric field creates a magnetic field.*
2. *Conversely, a moving magnetic field creates an electric field.*

It has been shown in an earlier chapter that the elementary electric generator depends for its operation upon the electric field created by a *moving* (relatively) magnetic field. When the conductors of the armature sweep by the magnetic field originating at the field poles, a difference of potential is induced in the conductor. This difference of potential constitutes an electric strain, or electric field, that is *always* at right angles to the magnetic field direction of movement Furthermore, this electric field exists whether there is a conductor present or not

It has also been pointed out that a moving electric field creates a magnetic field. In the usual case of a conductor carrying a current, the moving electric field may be considered as emanating from the countless moving electrons that make up the current flow. However, the presence of moving electrons and a conductor is not necessary, except indirectly, to create a magnetic field. For example, a magnetic field is created around the space between the plates of a capacitor connected to an alternating voltage, as discussed in Chap. VIII, although no current actually flows between the plates. The magnetic field is induced by the changing electric field between the plates. If the space between the plates of a condenser is enlarged, the field will be a comparatively open one, a field that corresponds to the field of an antenna.

**The Induction Field.** An elementary antenna system connected to a

source of h-f alternating current, such as a vacuum-tube oscillator, is illustrated in Fig. 233. The antenna consists of two wires, or legs (*AB* and *CD* in Fig. 233), and for the purpose of analogy, the two legs can be considered the plates of a capacitor. The a-c generator at the center is the source of energy and corresponds to the point at which the transmission line would be coupled in a transmitting antenna.

Since the legs of the antenna are in a vertical plane, the electric field is also in a vertical plane and is represented by the dashed lines in Fig. 233. The magnetic field created by the movement of the electric field, which is at right angles to it, is in a horizontal plane and is represented by the solid lines in Fig. 233.

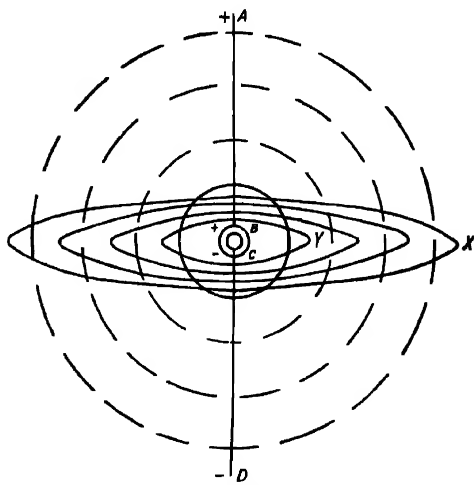


FIG. 233

It has been shown in the study of elementary electricity and magnetism that the strength of a magnetic field is proportional to the current flow in the conductor. It has also been shown in the study of capacitor action that, if a sinusoidal applied voltage is assumed, the charging current is greatest at the instant when the capacitor is totally discharged and is zero when the capacitor is fully charged.

In the capacitor formed by the legs of the antenna in Fig. 233, it is apparent that when the electric field is zero (*capacitor completely discharged*), the current is maximum and the magnetic field is fully expanded. Since one of the fields is at maximum when the other is at minimum, the electric and magnetic fields are 90° out of time phase, a fact that will presently be shown to be of vital importance at radio frequencies.

This combination of electric and magnetic fields in the immediate vicinity of a conductor and resulting directly from the current is called the **induction field.** Such a field will be found surrounding every a-c circuit, regardless of frequency. Although theoretically extending to infinity, the induction field at commercial frequencies is very limited, and practically all the energy is returned to the circuit upon the collapse of the field.

**The Radiation Field.** At greater distances from the antenna, a point is reached where the larger part of the electric field originates from the moving magnetic field rather than from the charge on the conductor,

and where the larger part of the magnetic field originates from the moving electric field rather than from the current in the conductor. This isolated combination of electric and magnetic fields constitutes the *radiation field*. This radiation field is self-propagating at a velocity that is so high that the moving electric field induces a magnetic field great enough at that velocity to sustain the electric field by induction. Thus, it is seen that the velocity of light and radio waves depends on the relation between electrostatic and electromagnetic induction. The velocity of radio waves is $3 \cdot 10^{10}$ cm per second, and the ratio between the electromagnetic and the electrostatic unit of charge is also $3 \cdot 10^{10}$.

In the induction field, the electric- and magnetic-field vectors are 90° out of phase, and most of the energy in the fields is returned to the conductor two times per cycle. In the radiation field, the electric and magnetic vectors are in time phase, and all the energy is lost to the conductor. The intensity of the induction field falls off much more rapidly with distance than does that of the radiation field. Since the area of the wave front of the radiated wave increases as the square of the distance from the antenna, the field energy per unit volume decreases as the square of the distance. Since the energy is proportional to the square of the field intensity. the intensity varies inversely as the first power of the distance from the radiator. At a distance greater than a few wave lengths from the antenna, the induction field is negligible, but the radiation field may continue to be appreciable for very great distances.

If the frequency is kept constant, the original strength of the radiation field will depend directly upon the magnitude of the current flow in the radiators (antenna conductors). In other words, an increase in the power input to the antenna results in an increase in antenna conductor charge and discharge current with consequent increased field strength.

From the foregoing it would appear that maximum radiation could be obtained by merely increasing power input and frequency. Unfortunately, a number of other factors, which impose very definite limitations, must be taken into consideration. Only a portion of the power absorbed by an antenna is radiated usefully in the form of an electromagnetic field. The remainder is consumed in various ways and, since it contributes nothing toward radiation, represents a complete loss.

## *ANTENNA RESISTANCE*

When dealing with power absorbed in a circuit, it is customary to regard this power as being expended in a resistance. The value of this resistance, of course, is such as would consume the actual power in the circuit at the same current flow. In antenna circuits, this fictitious resistance is known as the **effective resistance.** Since all the power expended

in an antenna is not usefully consumed as radiation, the effective resistance is divided into two components—*radiation resistance* (or useful resistance) and *loss resistance.*

Obviously, for maximum efficiency the loss resistance must be kept at a minimum. A number of factors contribute to the total loss resistance in an antenna system, and their importance should not be underestimated. The total loss resistance is composed of *dielectric loss*, *resistance loss*, *eddy-current loss*, *leakage loss*, and *corona loss*.

**Dielectric Loss.** Dielectric loss occurs in all dielectrics and varies with the frequency. It is due to the phenomenon of hysteresis and is especially noticeable in poor dielectrics, such as damp wood, concrete, masonry, trees, which are in the field of the antenna. Although this loss decreases as the frequency increases, it is one of the most important losses occurring in a radiating system and should be kept at a minimum even in h-f antenna systems. Special care should be taken also to keep poor dielectrics away from the ends of an antenna, since the highest electric gradients usually occur at these points. Much power loss due to poor dielectrics also occurs at the point of entry of antenna feeders into the transmitter room. If possible, a point of entry should be chosen where the material immediately surrounding the feeders has good dielectric properties.

**Resistance Loss.** Resistance loss is the easily understandable loss due to the actual ohmic resistance of the antenna and feeder conductors themselves. This loss can be kept at a minimum by making use of wire having a large cross section and good conductance. The largest useful cross section can be obtained by using a number of small wires properly interwoven and insulated from each other, thus minimizing the skin effect and reducing the r-f resistance. Tubular copper conductors are also often used for this purpose. In choosing the material for antenna wire, a happy medium must be reached at which good conductivity is combined with high tensile strength. Phosphor-bronze and silicon-bronze are often used.

**Eddy-Current Loss.** The emf induced in surrounding metallic objects within the antenna induction field causes eddy-current loss. Every effort should be made, so far as practicable, to eliminate all metallic masses from the vicinity of the antenna field. If the method of construction is such that guy wires are necessary for the support of antenna masts, such guy wires should be properly broken up by insulators, which should be so spaced that the intervening sections of guy wire will not resonate at any harmonic of the antenna frequency. If other antennas are in the vicinity, they should be as far removed from the transmitting antenna as possible. When it is necessary to erect a number of transmitting and/or receiving antennas in the same general vicinity, the radiating portions of antennas should be erected at right angles to near-by antennas in so

far as is possible. Eddy-current loss increases directly as the frequency and can assume troublesome proportions in h-f installations.

**Leakage Loss.** Leakage loss varies directly as the square of the voltage and inversely as the resistance of the leakage path. Leakage losses are caused by the flow of leakage currents from antenna to ground or between the legs of dipole antennas. The leakage-path resistance is naturally greatly diminished in wet weather, and the losses become noticeably large. Care should be taken to make all possible leakage paths as highly resistant as possible by the choice of proper insulators and, in some cases, feeder cables. Some of the peculiar shapes of commercial insulators are the result of special design to provide long leakage paths. Although leakage loss does not assume Herculean proportions on the higher frequencies, in the interests of efficiency every effort should be made to keep it at a minimum.

**Corona Loss.** Corona loss takes place at high voltages and is caused by partial ionization of the air about the antenna wires. The ionization causes the air to become a partial conductor and enables it to carry a current. Because of the bluish glow accompanying it, **corona effect,** as it is called, is visible at night. Corona effect begins to take place only at certain definite voltages, the critical voltage varying with the size and shape of the conductors and being smallest at points and corners in the antenna system. Corona loss varies inversely with the frequency, and on many of the higher frequencies it is practically negligible. All sharp turns and angles in the antenna conductors should be avoided if corona loss is to be kept at a minimum.

In a well-designed antenna system, the total loss resistance can be kept down to 5 per cent or less of the effective resistance if reasonable precautions are taken during installation and erection. In Chap. XVII methods of measuring the effective resistance as well as the inductance and capacitance of an antenna are discussed.

## *ANTENNA CIRCUITS*

The properties of an antenna used to receive radio waves are similar in nearly all respects to the corresponding properties of the same antenna when acting to radiate, or transmit, radio waves. As a matter of fact, the characteristics of a commercial receiving antenna are often determined by using the antenna as a radiator and making observations of the field strength with a portable receiver. The only difference between receiving and transmitting antennas is in the radiation-resistance component of the effective resistance. In the case of receiving antennas, the radiation resistance depends upon the load impedance coupled to the antenna. In view of the reciprocal relations between the transmitting and receiving properties of antennas, this discussion will not treat antennas separately

according to function but will be concerned primarily with transmitting antennas. Receiving antennas will be discussed only in so far as practical installation considerations and coupling methods differ.

Antennas can, in general, be divided into two types, namely, the Hertz, or doublet, antenna and the Marconi, or grounded, antenna. The former functions by virtue of the capacitance between two wires suspended in air and is used in modern transmitting systems mainly for h-f work. The latter functions by virtue of the capacitance between a single wire suspended in air and the earth. It is used widely for the lower frequencies, such as ship i-f transmission and broadcast work.

**The Hertz Antenna.** The Hertz antenna receives its name from Heinrich Hertz who used two short brass rods as radiators in his early

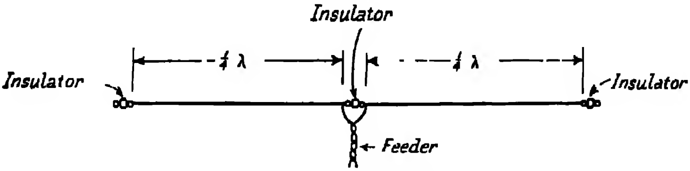



FIG. 234 The doublet antenna

experiments with electromagnetic induction. This arrangement of two conductors in space (see Fig. 234) has since been termed a **doublet.**

At low frequencies in general the current at different points of a circuit is the same. If the voltage remains constant, however, the current into a capacitive circuit increases with the frequency. Similarly, if the current remains constant, the voltage across an inductance increases with the frequency. At low frequencies, displacement currents are present only at those points of a circuit where relatively large capacitors have been intentionally inserted In other words, the inductance and capacitance are definitely localized, or lumped.

At very high frequencies, however, when the dimensions of the circuit become comparable to the wave length, the capacitances in different parts of a circuit become important, and the current varies appreciably in different parts of the circuit. This variation of current through inductances in different parts of the circuit causes their inductive effect to vary similarly. At high frequencies, therefore, the *equivalent* inductance and capacitance of the circuit will depend upon the frequency.

Such a circuit is shown in Fig. 235, where the circuit is composed of two long parallel lines supplied by an alternator at one end and closed at the far end. The inductance and capacitance of the circuit are represented by the dashed lines. At low frequencies, very little current flows through the circuit capacitors, and the reading of all the circuit ammeters would be essentially the same. As the frequency is increased, however, more and more current will flow through the capacitors because of their smaller

reactance. Consequently, the ammeter readings would decrease as the far end of the circuit is approached. The current reading at point $A$ (Fig. 235) would be highest. At point $B$ the current reading would be less by the amount of current that has been by-passed through the first capacitor. The current at point $C$ would be still lower and at point $D$ would be lowest of all. In other words, in such circuits one must deal with *distributed* inductance and capacitance.

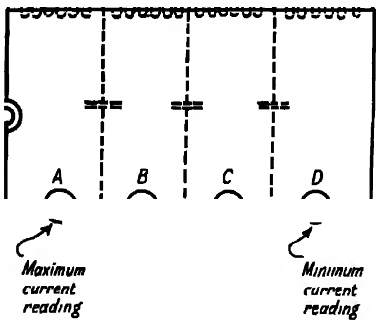



FIG. 235.

The distributed inductance and capacitance in antenna circuits are responsible for the voltage and current distribution in such circuits. Figure 236 shows the current-voltage distribution in a half-wave doublet antenna of the type shown in Fig. 234.

The mathematical treatment of currents in circuits having distributed inductance and capacitance is generally concerned with the case where these quantities are uniformly distributed. Because of end effects, this condition cannot be strictly realized in antenna circuits, but the wave forms of Fig. 236 can be considered essentially correct for most practical purposes.

The conditions that obtain in a simple antenna circuit can more easily be grasped by means of an elementary analogy. Consider a short length of rope lying on the ground. If one end of the rope is grasped in the hand and a quick vertical flip given to it, a wave, or loop, will travel down the rope to the far end. If both ends of the rope are grasped, one end in the left hand, the other in the right, a wave caused by a flip of the left hand can be made to travel along the intervening length of rope lying on the ground over to the right hand. In the same manner, a flip of the right hand will send a wave along the rope over to the left hand.

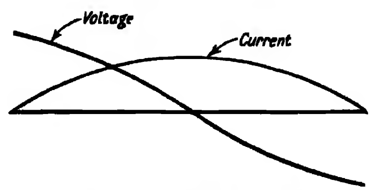



FIG. 236. Voltage and current distribution on a half-wave doublet antenna.

If the rope is short enough and the flips imparted to it do not follow one another in too rapid succession, alternate left- and right-hand flips, or pulses, can be sent along the rope in opposite directions without interfering with each other. This can be compared to applying an alternating current to a circuit having distributed constants.

In order to avoid mutual interference, the pulses must be so spaced that the traveling wave caused by a left-hand pulse reaches the right

hand before a right-hand pulse is imparted to the rope, and vice versa. It can be seen that the shorter the rope, the more easily and successfully this can be accomplished, and the longer the period between pulses (the lower the *frequency* of the pulses), the more successfully interference is avoided.

This condition may be compared to an a-c circuit at low frequencies. The frequency of the applied emf is analogous to the frequency of the pulsations applied to the rope, and the length of the rope may be compared to the physical dimensions of the electric-wire circuit when these dimensions do not approach the wave length of the applied emf.

If the rope in the analogy is made extremely short, it can be seen that no matter how rapidly the alternate left- and right-hand pulsations are applied, there is no interference between left- and right-hand waves in the rope. This is because the extreme shortness of the rope permits the waves formed by left-hand pulses to reach the right-hand end of the rope before the right-hand pulses are applied, and vice versa.

If the rope is made very long and the pulses are rapid enough, a number of waves can be made to travel through the rope simultaneously in different directions, with a hopelessly jumbled wave pattern in the rope. At certain critical lengths of the rope, however, for a given frequency of imparted alternated pulsations, a stationary wave can be made to appear on the rope. This is the result of a left- and a right-hand actuated wave traveling along the rope in opposite directions and reaching a given point on the rope with their peaks in phase, causing a stationary loop, or *standing wave*. Of course, since the rope is not rigid, the standing wave would collapse almost immediately after forming. If it were possible, however, to continue supplying alternate pulses to this rope and maintaining the amplitude and frequency of the pulses very accurately, enough energy would continuously be supplied to the rope to maintain the standing wave indefinitely.

The condition just described is analogous to a wire circuit when the frequency is very high and the physical dimensions of the circuit approach the wave length. Any simple wire circuit having distributed constants can be made to conform to this rule. If the frequency is increased to the point where the wave length approaches the dimensions of the wire, standing waves will appear on the wire. In any r-f circuits in which standing waves appear, radiation occurs. In l-f circuits, such as 60-c power circuits, a wire circuit would have to be several miles long in order to approach the wave length of the circuit and have standing waves. Under certain conditions, standing waves sometimes appear on long-distance power-transmission lines; but because of the low frequency, the radiated energy is small.

Since the purpose of antennas is to obtain as much useful radiation as possible, antenna design consists mainly of arranging a circuit having

distributed constants so that standing waves are produced. Because the feeder lines that are used for the transmission of excitation to antennas also carry r-f currents, standing waves sometimes appear on these lines if the lines are of the proper dimensions. Radiation from the feeder lines is undesirable in most installations. The elimination of standing waves on transmission lines is discussed in a later section of this chapter.

Further consideration of the foregoing rope analogy will show that standing waves cannot be produced when the rope is shorter than ½ wave length (one loop). Similarly, an ungrounded antenna cannot be resonated (hence, excited for radiation) if its electrical length is less than a half wave length. This statement is further qualified for grounded antenna systems in the section on the Marconi antenna.

Further experimentation with the rope would show that as the length is increased from a ½ wave length, additional standing waves could be formed only at certain critical lengths of the rope, the frequency of pulsation remaining constant. These critical lengths would be found to be any multiple of the original ½ wave length. The same holds true for antenna circuits. In ungrounded antennas, resonance is obtained whenever the electrical length of the antenna is a multiple of a half wave length. Thus, if it were desired to radiate a 3,000-kc signal (100 m), an antenna ½ wave length long (50 m) could effectively be coupled to the transmitter, an antenna 1 wave length long (2 · 50, or 100 m) could be used, an antenna $\frac{3}{2}$ wave lengths long (150 m) could be used, and so on.

An antenna, as a result of its distributed inductance and capacitance, acts in many respects as a resonant circuit. Consequently, when an antenna is adjusted to exact resonance with the transmitter that is exciting it, it presents a resistance load. If the frequency of the voltage applied to the antenna is varied above and below the resonant frequency of the antenna, capacitive reactance is presented by the antenna at frequencies just below resonance and inductive reactance at frequencies above resonance. Therefore, if an antenna is not exactly the correct length for resonance, it can be compensated by adding series inductance or capacitance. If the antenna is too short, series inductance can be added to neutralize the capacitive reactance of the antenna and bring it into tune. Similarly, capacitance can be added to neutralize the inductive reactance of an antenna that is too long.

The wave forms for a number of simple antennas of different lengths are shown in Fig. 237. It should be noted that minimum current and maximum voltage always occur at the ends of an antenna, regardless of the length of the antenna. Points of maximum amplitude (either voltage or current) in antenna circuits are referred to as current, or voltage, **loops**. Points of minimum amplitude are called voltage, or current, **nodes**.

The theoretical treatment of radiating antennas usually concerns an antenna in free space. Actually, of course, all antennas are comparatively

close to the ground and are therefore subject to the reflection characteristics of the earth. In any antenna system, a portion of the radiated waves are directed toward the earth at various angles and are reflected skyward again with varying effectiveness, depending upon the conductivity of the earth. Perfect reflection results only if the earth is a perfect conductor (zero resistance). Since the actual conductivity varies

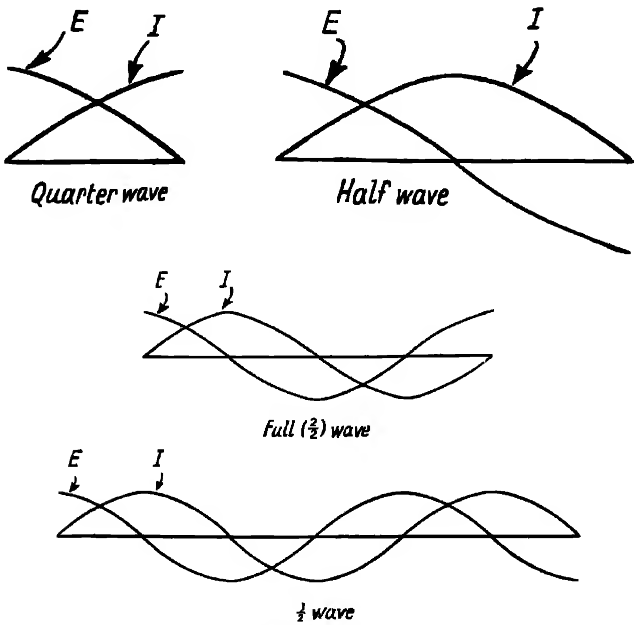



FIG. 237 Voltage and current distribution on simple antennas.

greatly, it is difficult to calculate with any degree of success the actual reflection characteristics of the earth. Perfect conductivity can be assumed for most cases with negligible practical error in computing the efficiency of an antenna. With perfect conductivity of the earth, the field produced is the same as would be produced with the earth removed and with the aid of an additional antenna located a distance below the former surface that is equal to the distance of the actual antenna above the surface. This fictitious antenna would have characteristics identical with those of the actual antenna and is often called the **image antenna.** Image antennas are included in the calculations involved in predicting the field of an antenna.

**The Marconi Antenna.** Marconi, or grounded, antenna systems are customarily used on the lower frequencies. Ship i-f (500 to 375 kc) installations and broadcast-band (550 to 1,600 kc) installations are customarily of this type. The physical dimensions of ungrounded antennas

become quite formidable at low frequencies, even at minimum length ($\frac{1}{2}$ wave length), and their use for l-f transmission is therefore comparatively limited.

The use of grounded antenna systems permits utilization of radiators of less than $\frac{1}{2}$ wave length. This arrangement does not conflict with the theory previously discussed for ungrounded antennas. The smallest self-resonant ungrounded antenna is an electrical half wave in length. If such a half-wave antenna were cut in two, making its length $\frac{1}{4}$ wave length, and if one end were grounded, the system would still function as

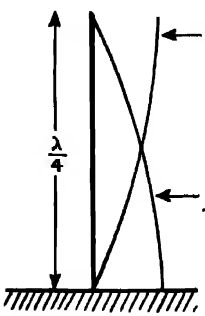



Fig. 238 Voltage and current distribution on a grounded quarter wave vertical radiator

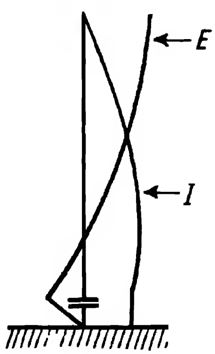



Fig. 239

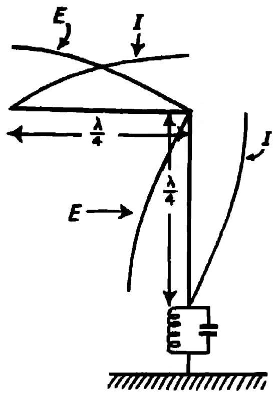



Fig. 240

a half-wave radiator. In other words a grounded quarter-wave antenna will resonate at the same frequency as an ungrounded half-wave antenna. The voltage and current distribution in a quarter wave antenna is the same as the distribution in half of an ungrounded half-wave antenna and is shown in Fig 238. The missing half of the antenna can be considered as being supplied by the image antenna in the ground.

In common with ungrounded systems, a grounded antenna may be much smaller than a quarter wave and still be made to resonate at the desired frequency by "loading" it with series inductance. However, the efficiency of the antenna as a radiator is materially reduced, since the highest current in the circuit is at the point where the inductance must usually, perforce, be introduced (the part of the antenna nearest the ground). Radiation from the inductance itself is practically negligible.

A large portion of the loss resistance in grounded antenna systems occurs in the ground contact. The current is at a loop at this point. A greater ratio of radiation resistance to loss resistance can be obtained by decreasing the antenna current at the ground contact point, which can be accomplished by increasing the antenna length, with a resulting shift in current loop as shown in Fig. 239. With such an arrangement

greater antenna efficiency is realized for a given power input to the antenna, since a larger portion of the radiating wire carries greater amplitude of current. The antenna circuit is brought back into resonance by inserting a series capacitance. Maximum value of radiation resistance

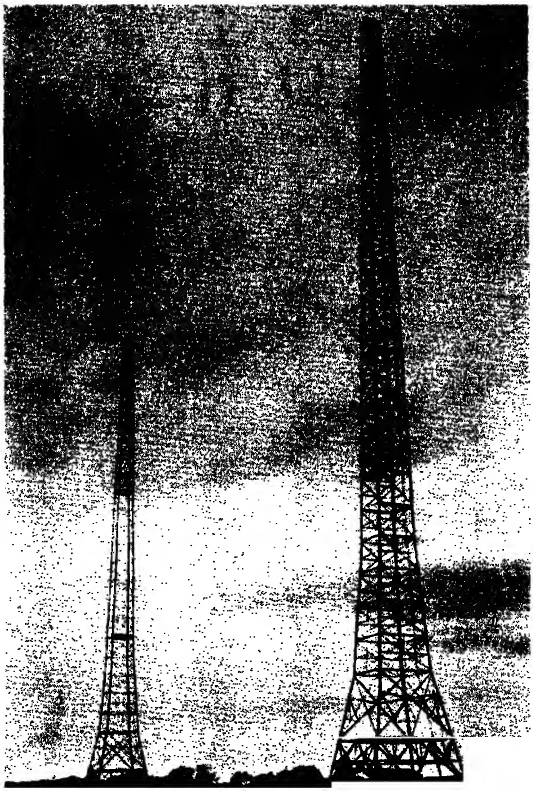


FIG. 241. Self-supporting towers of station WNBC, Port Washington, N.Y. Note buildings housing the antenna coupling units and transmission line in the foreground. (*Courtesy of Lehigh Structural Steel Co.*)

occurs when the antenna length is $\frac{1}{2}$ wave length, since this length brings a current node at the ground connection.

The inverted-L and T-type antennas often found on shipboard are common forms of quarter-wave grounded antennas. The hull of the ship and the surrounding water form a very efficient ground system. In such antenna systems, the radiating portion of the antenna is partly vertical

and partly horizontal, with resulting current and voltage distribution as shown in Fig. 240.

Nearly all modern broadcast antennas utilize self-supporting steel towers. The towers themselves serve as vertical radiators and have a height either approximately $\frac{1}{4}$ wave length or slightly in excess of $\frac{1}{2}$ wave length. Such antennas are customarily excited by insulating the base of the tower from ground and applying the exciting voltage between the tower base and ground. Other systems utilize a grounded tower. The feeder wire is attached to the tower at a height corresponding to approximately 20 per cent of the tower height, as shown in Fig. 242(a). Excitation current flows through the loop formed by the lower section of the tower, the ground, the transmitter output circuit, and the feeder wire.

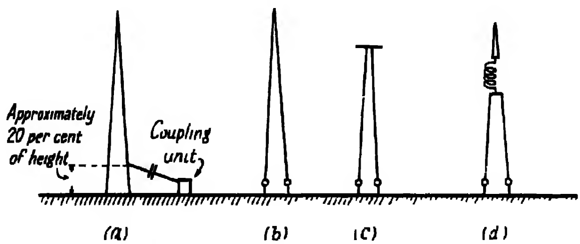



FIG. 242. (a) Grounded tower radiator (b) Simple unloaded tower radiator (c) Tower radiator with capacity flat top (d) Tower radiator with series inductance loading

Sufficient voltage develops across the lower section of the tower to excite the remainder of the tower. The loop circuit is resonated by means of a capacitance in series with the feeder line

The relatively great cost of tower radiators of optimum height has led to the development of several alternative arrangements. One system utilizes a tower having a large flat top. The large capacitance of such a top increases the natural electrical period of the tower, so that a quarter-wave flat-top tower for a given frequency is shorter than an ordinary quarter wave tower for the same frequency. The saving in cost in practical flat top tower construction is so small as to be a minor consideration, however. Another scheme is to break the tower into sections, inserting series inductance between sections. Although sectionalized towers have been constructed with series inductance that permitted reductions in tower height of 20 to 30 per cent, this saving is largely offset by the extra cost of sectionalizing and by the losses in the inductance. Sectionalized towers, however, are often used where the installation is subject to experimentation, for they permit experimental adjustments of effective tower height without structural changes.

As in other grounded-antenna systems, one of the major losses in tower installations is that occasioned by the ground-contact resistance. Broadcast stations considerably reduce this loss by providing a ground

system of buried wires, as shown in Fig. 244. Such a ground system consists of a number of radial wires directly underneath the radiating tower. Optimum specifications of a ground system call for radials having a minimum length of 0.4 wave length and spaced about 3° apart. Since

FIG. 243 Tower base insulator The unit shown is capable of supporting a weight of 3,000,000 lb (*Courtesy of Lapp Insulator Co., Inc.*)

maximum ground loss occurs in the immediate vicinity of the tower base, most installations utilize an additional local ground system extending a short distance from the tower base. Such ground mats are often composed of a buried wire screen.

**General Antenna Considerations.** The direction of the electric field radiated by an antenna is also the direction of polarization of the wave. Thus, a wave with its electric field vertical is said to be **vertically polarized**; one with a horizontal electric field is said to be **horizontally polarized.**

The antennas that produce these fields are classified according to the polarization of the field. Thus, a vertically polarized antenna is one that produces a vertically polarized electric field, and, similarly, a horizontally polarized antenna is one that produces a horizontally polarized field. In general, the polarization of an antenna depends upon its position with respect to the earth's surface. Thus, a vertical antenna is vertically polarized and a horizontal antenna horizontally polarized.

Low-frequency radiated waves usually maintain their original polarization, but as the frequency is increased, it is found that the polarization

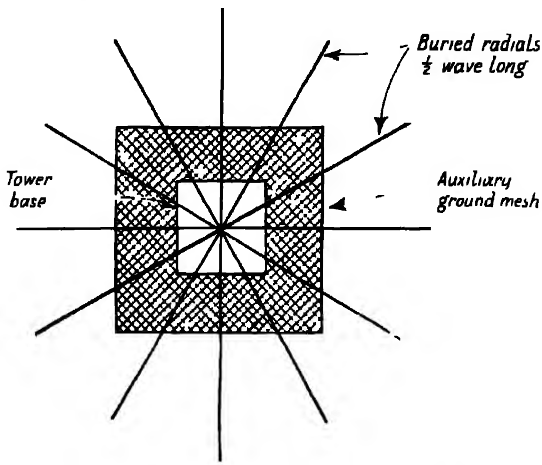



FIG. 244. Ground system for vertical radiator

changes during travel. The radiated wave usually splits up into several components, which follow different paths in arriving at a given receiving antenna. As a result, the received wave is usually found to contain both vertical and horizontal components of polarization.

Horizontal polarization is usually preferred for transmitting antennas because the antenna-supporting structures required are not so high as for vertically polarized radiators. Since most radiated noise is vertically polarized, horizontally polarized receiving antennas are productive of a better signal-to-noise ratio and are therefore to be preferred.

The useful component of an antenna field has been shown to be the radiation field. Neglecting directivity effects, the waves radiated from an antenna under normal conditions travel in all directions from the antenna. It is apparent that a great portion of the radiated energy is lost in a non-directional system. The useful radiation is divided into two parts - the *ground wave* and the *sky wave*. The ground wave represents that part of the radiated signal which follows the general contour of the earth. The sky wave represents that part of the radiated signal which is

radiated skyward and is usefully reflected back to the earth by the ionosphere. This reflection characteristic is discussed in detail in a later section of this chapter dealing with the ionosphere.

Practically all commercial point-to-point radio transmission is accomplished by means of the sky wave, since the ground wave is appreciably

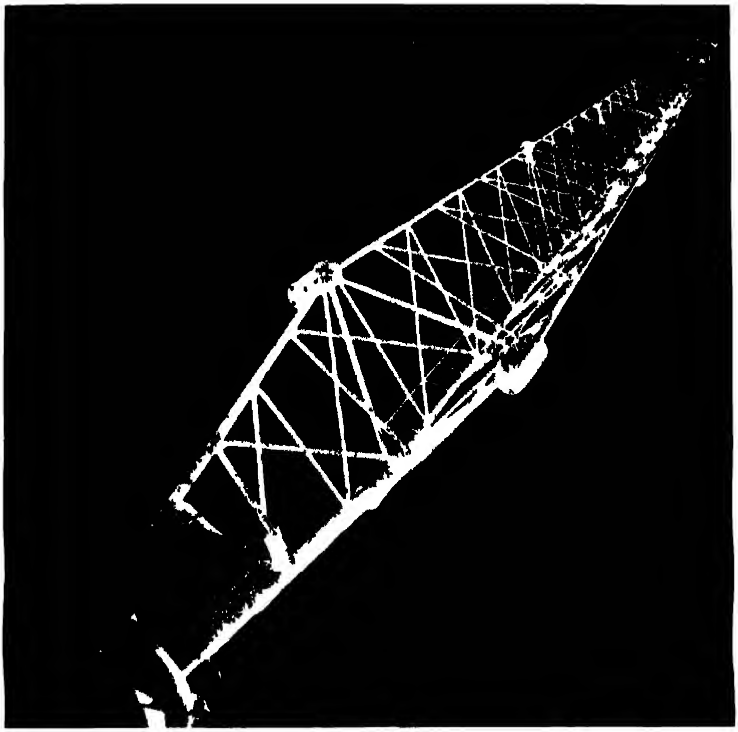


FIG. 245. 650 ft insulated half wave radiator of KMJ, Fresno, California. (*Courtesy of International Derrick and Equipment Co.*)

attenuated a short distance from the transmitter. Broadcast stations, the primary object of which is to deliver a high-quality signal to listeners within a limited area around the transmitter, depend upon the ground wave for primary coverage. Within this primary-service area, broadcast signals show little or no seasonal or other variation and have good entertainment value because of the high signal-to-noise ratio. Beyond the primary-service area of a broadcast station, there is a secondary-service area where, although signals are fairly strong, reception is less satisfactory than in the primary area. In the daytime, broadcast-frequency sky waves are almost completely attenuated. Secondary coverage is due

practically entirely to ground wave reception, with the result that the daytime secondary service area extends only a short distance beyond

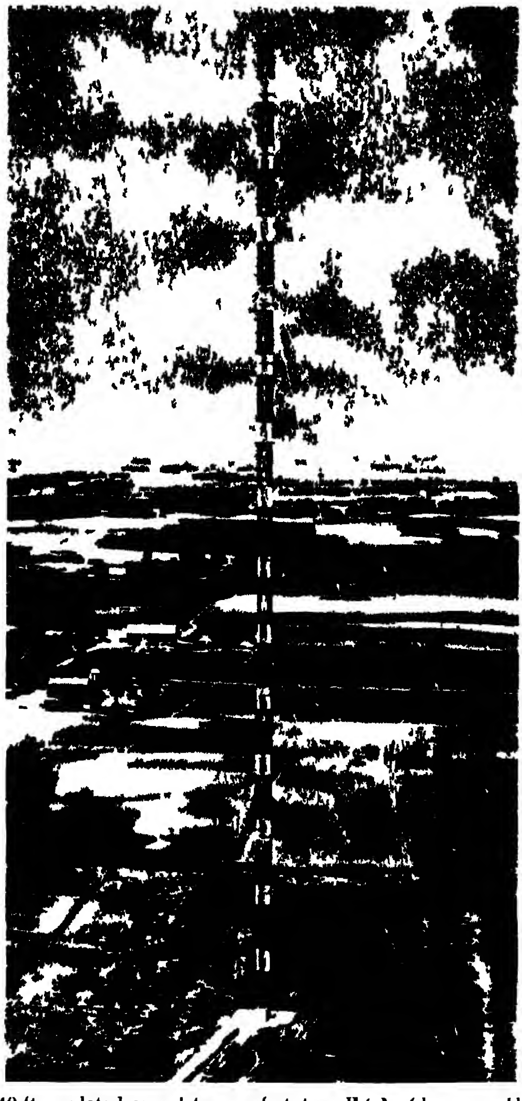


FIG. 246. 740 ft insulated guyed tower of station WGN, Chicago. The tower height corresponds to 190° of wave length or slightly more than ½ wave. (*Courtesy of Truscon Steel Co.*)

the primary service area. At night the sky wave is sufficiently strong to give good secondary coverage over large areas.

**Directional-antenna Characteristics.** Only the waves radiated in certain directions from a transmitter will reach a particular receiving point; all waves radiated in other directions are wasted so far as this receiver is concerned. For radio communication between fixed points, therefore, greater efficiency and lower power consumption are attained by the use of directional antennas. Directional antennas are also important for receiving purposes because signals arriving from directions other than the desired signal are attenuated, as are also static and man-made noise arriving from other directions. The result is a greater desired signal-to-interference ratio and signal-to-noise ratio. In addition, directional antennas for fixed service can be adjusted to resonate only at the desired signal frequency, resulting in considerable gain in the antenna system itself.

FIG. 247. Tapered self-supporting 640-ft tower of the 50-kw Ethiopian government station at Turi, Ethiopia. The tower height corresponds to 0.47 wave length. (*Courtesy of Truscon Steel Co.*)

The directional characteristic of a single-wire antenna in free space is broadside to the wire, as shown in Fig. 248(a). This polar diagram of a space pattern is for a single-wire antenna ½ wave length long. As the length of the antenna is increased by multiples of ½ wave length, it is found that the two original lobes of the field pattern are multiplied according to the length of the antenna. Thus an antenna 1 wave length long has four lobes, one in each quadrant, all of approximately equal amplitude. As the antenna length is further increased, it is found that the number of lobes in the field

pattern increases. There are as many lobes in each quadrant of the field pattern as there are wave lengths in the antenna length. Thus, the antenna 1 wave length long shown in Fig. 248(b) has one lobe in each quadrant. An antenna 5 wave lengths long (Fig. 248(c) ) has five lobes in each quadrant.

As the antenna is increased over one wave length, the lobes in the field pattern do not remain equal in amplitude, since one lobe in each

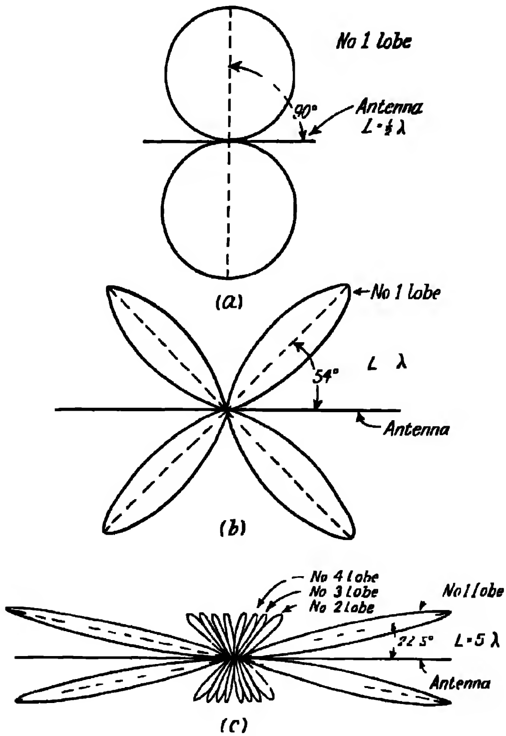



Fig. 248

quadrant always predominates. As the antenna length is increased, the angle that the major lobe in each quadrant makes with the antenna wire becomes smaller and smaller, with the result that very long single-wire antennas have marked directional characteristics in the direction in which the wire is pointing.

The directional characteristic of a long-wire antenna differs from that of an elementary half-wave antenna because the current in different parts of the long antenna may not always be in the same phase. Furthermore, the distance from a remote receiving point to various parts of a long antenna will not be the same. Consequently, the fields radiated from

different parts of a long antenna add together vectorially and the net field depends upon the direction from the antenna and upon the current distribution in the wire.

A vertical antenna is essentially nondirectional. Thus the horizontal-plane field pattern of a vertical antenna is a circle; that is, looking down on top of the antenna in such a way that the antenna wire appears as a dot, the field pattern will appear as a circle. Nevertheless, combinations of vertical antennas in spaced arrays are used with great success to obtain unidirectional characteristics. Any directional antenna system involving two or more spaced antennas is termed an **antenna array.**

One of the simplest forms of antenna arrays consists of two vertical antennas spaced $\frac{1}{4}$ wave length apart and excited by equal currents 90° out of phase. In one direction along the plane of both antennas the phase relations of the two antenna fields are such that the fields add together and result in maximum field strength in this direction. In the opposite direction along the same plane, the field phase relations are such that the fields balance and give zero field in this direction. In all other directions from the antennas, the net field is the vector sum of the two antenna fields, and the field strength is always less than the field strength in the direction of maximum amplitude. The resulting horizontal plane wave pattern is a cardioid, as shown in Fig. 249.

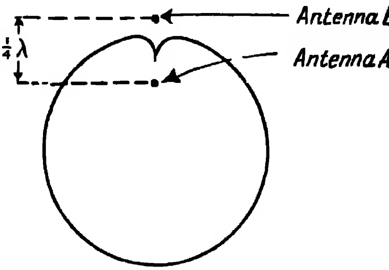



FIG. 249. Resulting field pattern when two vertical radiators, spaced $\frac{1}{4}$ wave length apart, are excited 90° out of phase.

Practically all the numerous types of antenna arrays, including broadside, end-fire, and collinear arrays, are adaptations of the fundamental principles discussed for the two-element array above. Some commercial installations even utilize arrays of arrays.

The individual antennas making up an array may be of any desired type. Thus, it is possible to use horizontal, instead of vertical, radiators. Radiators of any length ranging from $\frac{1}{2}$ wave length on up may be used. Of course, all the elements in a given array should be of the same length.

The simplest form of **broadside array** consists of a number of antennas uniformly spaced along a line, all of which are excited by equal currents of the same phase. Such an array results in a concentration of radiation in a plane at right angles to the line of the array. The field in directions other than at right angles to the array line is the vector sum of the individual-element fields and is almost all canceled. The sharpness of the field pattern in the direction of greatest amplitude is a function of the length of the array, increasing with array length. The individual-element

spacing must be equal between elements, but otherwise it has a negligible effect on the field pattern up to a certain critical distance. The critical spacing varies with particular systems and depends upon the length of

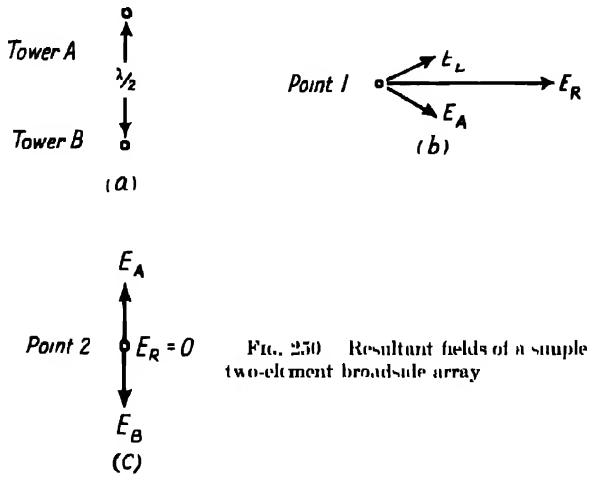



FIG. 250 Resultant fields of a simple two-element broadside array

the individual elements, the number of elements in the array, and the operating frequency.

The operation of a broadside array can be more clearly understood by referring to the simple two element array of Fig. 250. The two vertical elements $A$ and $B$ are spaced ½ wave length apart and are fed by in-phase currents. At point 1, the voltage $E_A$, due to the wave from antenna element $A$, and the voltage $E_B$, from antenna element $B$, arrive in time

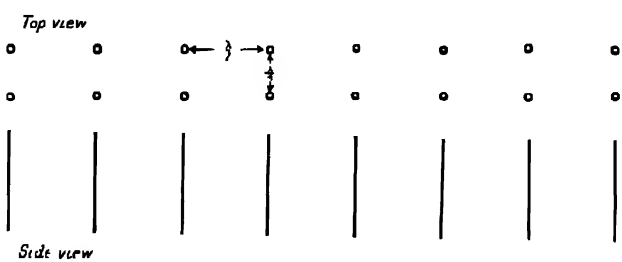



FIG. 251. A typical broadside-array arrangement

phase and, consequently, add to each other, making the resultant voltage $E_R$ maximum at this point. At point 2, voltage $E_A$ arrives in time-phase opposition to voltage $E_B$ since the $E_A$ wave had to traverse a path that is longer by one half wave length (the spacing between towers) than the $E_B$ wave path. Consequently, $E_A$ lags $E_B$ by 180° upon its arrival at point 2. The two voltages, being equal in amplitude and opposite in

polarity, cancel out, and the resultant voltage $E_R$ becomes zero. At intervening sections of the quadrant between points 1 and 2, the resultant field voltage assumes some value between the maximum of point 1 and the minimum of point 2.

A widely used form of broadside array consists of an arrangement of two-element arrays of the type discussed on page 414 and shown in Fig. 251. The broadside array consists of a number of couplets spaced $\frac{1}{2}$ wave length apart. Each couplet consists of two half-wave elements spaced $\frac{1}{4}$ wave length apart and excited by equal currents 90° out of phase. This results in a remarkably efficient unidirectional system. The field pattern of each couplet is similar to that shown in the horizontal plane polar diagram of Fig. 249 The horizontal-plane field pattern for the entire array is shown in Fig. 252.

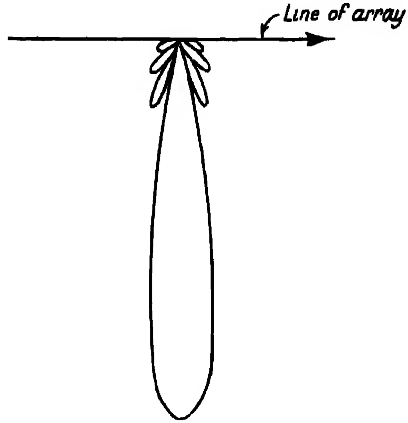



FIG. 252. Horizontal-plane field pattern of array shown in Fig 251

An **end-fire array** consists of a series of antennas arranged in a line and excited by currents of equal amplitude. The spacing in wave lengths between antennas is equal to the progressive phase difference in cycles in the currents between adjacent antennas. Thus, if the antennas are excited progressively 90° out of phase, a spacing of $\frac{1}{4}$ wave length is required. If the antennas are excited progressively 180° out of phase, a spacing of $\frac{1}{2}$ wave length is required. Optimum results are usually

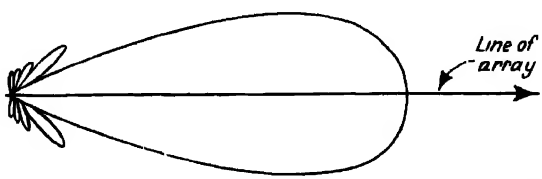



FIG. 253. Horizontal plane field pattern of typical end fire array

obtained with quarter-wave or three eighth-wave spacing between elements. End-fire arrays are relatively efficient. The horizontal-plane field pattern of a typical end-fire array is shown in the polar diagram of Fig. 253. The direction of radiation concentration is always toward the end of the array having the lagging phase.

As pointed out in an earlier section of this chapter, the lobe of maximum radiation from a single long wire makes a more acute angle with the wire

and increases in amplitude as the length of the wire in wave lengths is increased. There are a number of ways in which long single-wire antennas can be combined to increase the gain and directivity of an antenna system. One of the simpler forms of such a system is the so-called V antenna.

If two long single wires are combined to form a V, as shown in Fig. 254, a very effective bidirectional antenna results. When the major angle of the V is twice the angle that the major lobe of each wire makes with the wire, and the two sides of the V are excited 180° out of phase

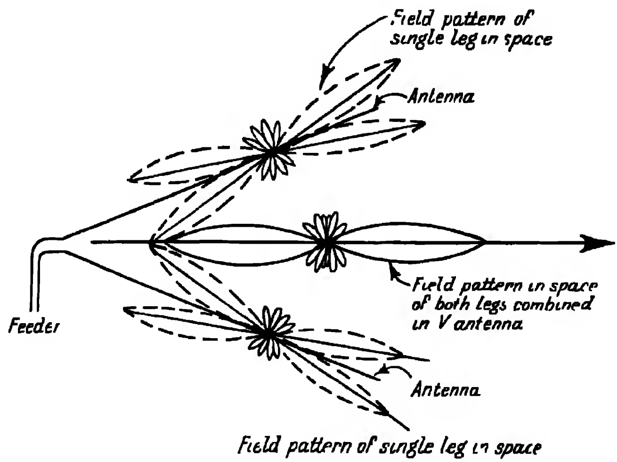



FIG. 254. V-type antenna showing horizontal-plane field patterns of individual and combined legs.

by connecting a two-wire feeder to the apex of the V, the major lobes reinforce along the line bisecting the V. Maximum radiation is concentrated along the bisector in either direction, and the lobes tend to cancel in all other directions. The horizontal-plane field pattern is shown in Fig. 254. The normal field pattern for each leg of the V in space is indicated by the dashed lines, and the field pattern for the V arrangement is shown in solid lines.

The wire lengths in a V beam are not at all critical, but it is important that both wires be of the same electrical length. The gain obtained depends upon the wire length in wave lengths.

The V antenna can be made unidirectional and aperiodic by terminating the open ends of the V to ground through resistors. These resistors dissipate almost half the power fed to the antenna (power that would be radiated in the reverse direction in a bidirectional V). Because of the losses occurring as a result of ground contact resistance, terminated-V antennas are not often used.

One of the most widely used directional antennas today is the **rhombic,**

or **diamond, antenna,** so called because a horizontal plan view of the antenna is a rhomboid, or diamond-shaped. The rhombic antenna differs somewhat in principle from the antennas previously described (with the exception of the aperiodic V antenna) in that it radiates without the presence of standing waves of current and voltage along the wires.

Fundamentally, the rhombic antenna consists of a pair of unidirectional V antennas placed side by side. The arrangement is shown in Fig. 255. The antenna is made nonresonant by terminating the far end of the diamond in a resistance that is equal to the characteristic impedance of

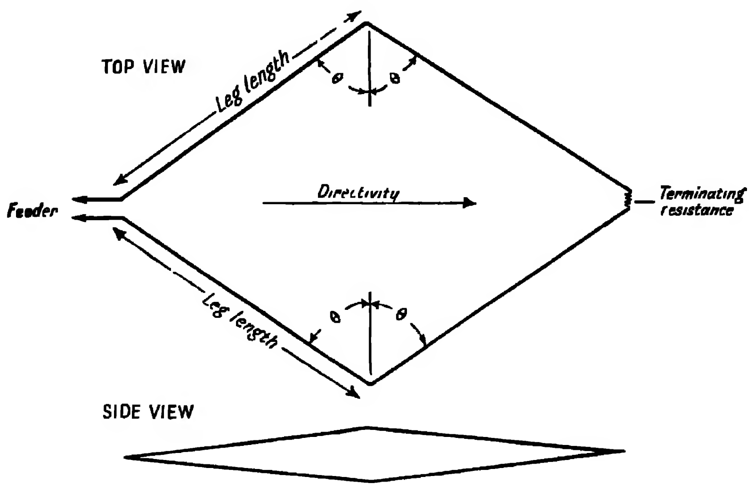



FIG. 255 The rhombic (diamond) antenna

the antenna. For a discussion of characteristic impedance the reader is referred to a later section of this chapter devoted to transmission lines. This terminating arrangement avoids the ground losses experienced with the terminated-V type antenna.

The major lobes of the individual wires forming the four sides of a rhombic antenna reinforce each other to give a directivity, as indicated in Fig. 255. If the terminating resistance were eliminated, the antenna would be bidirectional along this same line. The terminating resistance has the effect of eliminating standing waves on the antenna. The current in the antenna decreases uniformly along the wires as the terminated end is approached, the decrease being caused by loss of energy from radiation and by other normal antenna losses. The energy remaining when the terminus is reached is dissipated in the terminating resistor instead of being reflected back along the wire to form standing waves. The terminating resistor therefore absorbs all the power that would otherwise be radiated in the backward direction.

The directional characteristics of a rhombic antenna are not particularly critical with respect to the tilt angle (see Fig. 256), and, consequently, the antenna can be used over a considerable frequency range without adjustment. Optimum results are obtained when the length of each leg is not less than 2 wave lengths. A change in frequency has the effect of altering the lengths of the legs as measured in wave lengths, so that the net result of an increase in frequency is to sharpen the directional-field pattern somewhat materially changing its fundamental character.

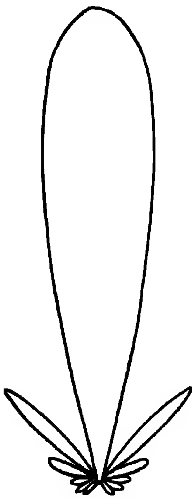


Fig. 256. Polar diagram showing horizontal-plane field pattern of rhombic antenna having a leg length of 4 wave lengths.

In practice, the sides of a rhombic antenna are composed of a pair of conductors in parallel with a variable spacing between them, as shown in Fig. 255. This arrangement offsets the variation in distributed inductance and capacitance caused by the unequal spacing of the two sides of the diamond. By virtue of this compensation, the characteristic impedance is essentially the same at all points along the antenna.

Several variations of design are available to produce optimum results with rhombic antennas. One of the factors entering into the design of almost all commercial antenna systems is the radiation angle, or wave angle. The **wave angle,** briefly, is the angle of maximum radiation above the ground for a specific antenna under consideration. This topic is fully discussed in the following section, and the method of computing optimum wave angle for a given fixed point-to-point transmitting or receiving antenna system is described.

The first step in the design of a rhombic antenna is to ascertain the optimum wave angle for the frequency and range to be used. Once the wave angle is selected, there are three remaining quantities to be determined: the tilt angle $\theta$, the antenna height $H$, and the length of each leg $L$.

For any given wave angle, there is one set of antenna constants that will produce maximum radiation in the desired direction or maximum response to signals coming from the desired direction. The optimum height of a rhombic antenna for a given frequency and wave angle is derived from the equation

$$H = \frac{\lambda}{4 \sin w} \qquad (1)$$

where $H$ = antenna height in meters;
$\lambda$ = operating wave length in meters;
$w$ = wave angle.

The tilt angle is determined directly from the relation

$$\sin \theta = \cos w, \tag{2}$$

where $\theta$ = tilt angle and $w$ = wave angle.

The leg length (for each of the four sides) is obtained from the formula

$$L = \frac{\lambda}{2 \sin^2 w}, \tag{3}$$

where $L$ = length of each leg in meters;
$\lambda$ = operating wave length in meters,
$w$ = wave angle.

**Aircraft Antennas.** Refraction of a radio wave occurs when the wave passes over a metallic sheet, such as the wing of an airplane. Receiving

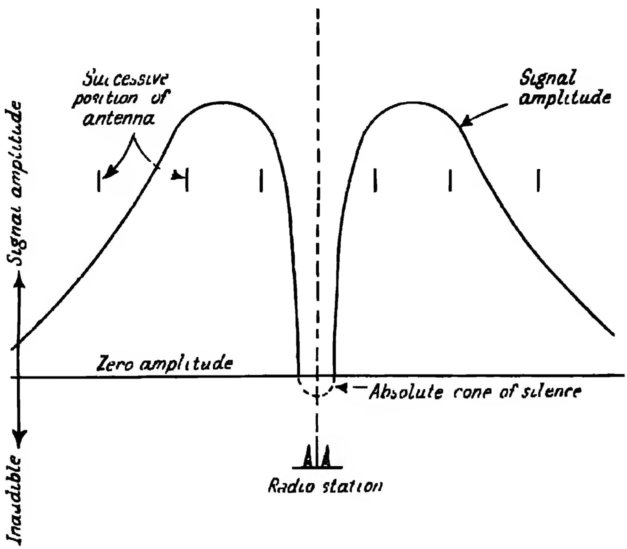



FIG. 257. Behavior of a simple antenna in space, which passes from left to right over a radio beacon station.

antennas located above the wing, but shielded by the wing from the transmitting station, will receive energy refraction over the edge of the wing. The refraction process tends to change the wave front of the wave and rotates the electric and magnetic-field components of the radiated field. When refraction takes place over both the leading and trailing edges of the wing, the wave front will be tipped, and partial cancellation may result for an antenna located above such a wing. This action is reciprocal in the case of transmitting from an airplane when the antenna is located above the wings.

Figure 257 shows the behavior of a simple vertical antenna in space

that passes from the left to right over a transmitting station. The change in signal amplitude is shown by the graphical representation. The antenna will show a *cone of silence* when passing directly over the transmitting station owing to the fact that waves traveling over different paths arrive at the antenna 180° out of phase. When the antenna passes over a beacon, or directional, transmitting station, the cone of silence will be *absolute*, or complete, over the indicated area. For other types of transmitting

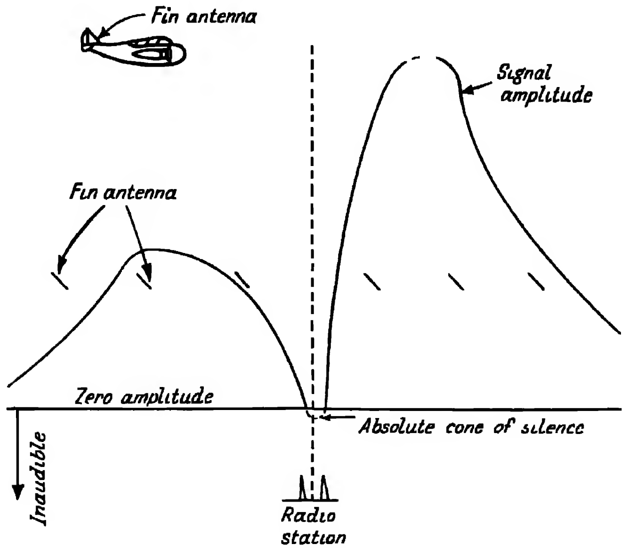



FIG. 258 Behavior of a standard fin antenna that is passing from left to right over a radio-beacon station.

stations, the cone of silence will be only partial, or incomplete, that is, complete cancellation of the signal will not occur.

A standard *fin antenna*, such as is used on the Douglas-type transport planes, is shown in Fig. 258. The antenna extends forward and downward to the top of the fuselage where the lead-in to the radio receiver is attached. As the plane moves toward the transmitting station, its projected angle toward the station decreases, and the normal surge of signal on approach is not so great as with the ideal vertical antenna. The signal strength builds up at a rapid rate, but the net result is that the surge is decreased in amplitude, and the decrease in signal beyond the surge does not have so great a slope as with the ideal antenna in Fig. 257. As the plane reaches a point over the transmitting station, the signal drops to inaudibility at a point almost exactly over the station. Complete inaudibility actually occurs only over a beacon transmitting station. Immediately after passing the station, the antenna presents a large projected area, and a surge is heard that is much greater than the surge on approach. The

rate of surge is extremely rapid, for the same reason that the decrease of signal from the surge before the station occurred at a rate slower than normal.

Figure 259 shows a *whip antenna* mounted below a plane that maintains an angle in flight of approximately the amount shown. This bending of the antenna causes the signal to increase on the first surge on approach to the transmitter, by about the same amount as on the surge that follows

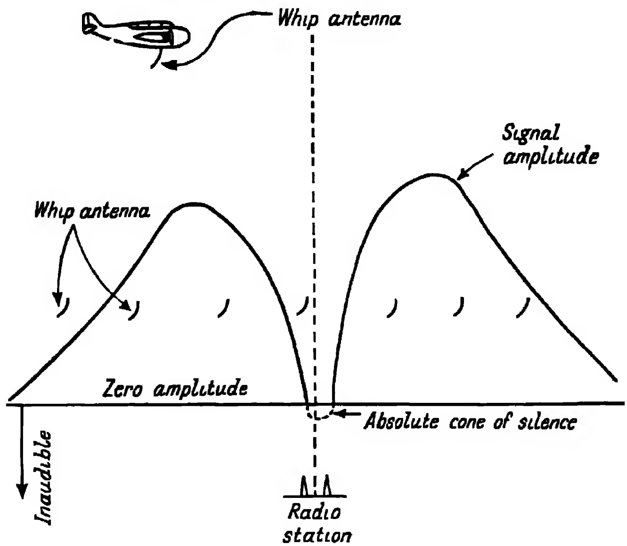



FIG. 259 Behavior of a whip antenna mounted below the aircraft, which is passing from left to right over a radio beacon station This type of antenna and installation most closely approaches the ideal of Fig 257

the passing of the station. The fact that the antenna has slightly better pickup on entering the cone than upon leaving it causes the decrease to the absolute cone from the first surge to be somewhat lengthened over that with the ideal antenna. It also causes the second surge to diminish less rapidly upon leaving the station. The absolute cone of silence is sustained for a longer time than with the fin antenna, because this antenna has less horizontal polarization component. At low altitudes (300 ft) over the transmitting station, the cone is very distinct and well-marked. This is the most effective type of all the antennas discussed.

The same whip antenna mounted on top of the plane is shown in Fig. 260. As the plane approaches the station the antenna, because of its bend, presents an increasingly smaller projected area to the station. This prolongs the decrease time into the absolute cone appreciably. As the plane comes into a position above the station, the signal is arriving from well below the plane and there are two paths of signal, one around

the leading edge of the wing and the other around the trailing edge. Considerable refraction takes place. and the wave front is tipped as it comes over the leading edge of the wing. At a point beyond the position directly over the station, the wing center is directly between the antenna on top of the plane and the transmitting station. Owing to this distortion of the wave front, the two paths of signal over the leading and trailing edges of the wing may give a minimum signal in this position. In actual

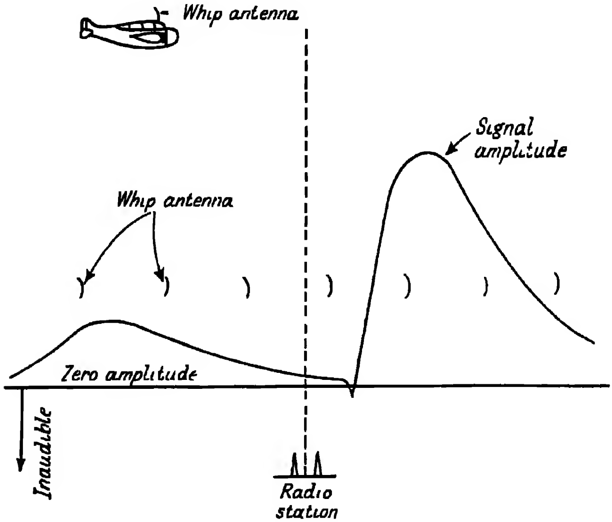



Fig. 260. Behavior of a whip antenna mounted above the aircraft, which is passing from left to right above a radio-beacon station

tests, the cone of silence has been found to occur at an angle from the vertical of 34° beyond the position over the transmitting station.

As the airplane approaches the station, there is a surge that is not so great as the surge following the station. A very gradual decrease of signal into the cone of silence results, which, as explained above, occurs approximately 34° beyond the vertical. This decrease is immediately followed by a very sharp rise in signal and heavy surge as the refraction effect disappears and the signal strikes the largest projected area of the antenna beyond the station.

Some interesting effects of whip antennas mounted in the positions discussed above are shown in Fig. 261 when the plane is circled at some distance from a beacon station. At (a) the plane is shown with the antenna mounted on top of the fuselage, and the plane is in a bank. It would be expected that, as the wing reaches a position where it is between the antenna and the beacon station, refraction over the wing edges would result in a sharp decrease in signal. This decrease is very pronounced

when circling at a distance of 15 or 20 miles from a beacon station. When the plane has circled 180°, some slight decrease in signal due to smaller projected area of the antenna toward the station is noted. This reduction in signal is not pronounced. As the bank is increased, both of the effects mentioned become more noticeable. A turn made with no bank at all produces negligible change in signal.

In Fig. 261(b), the effect of a bank with the antenna mounted below

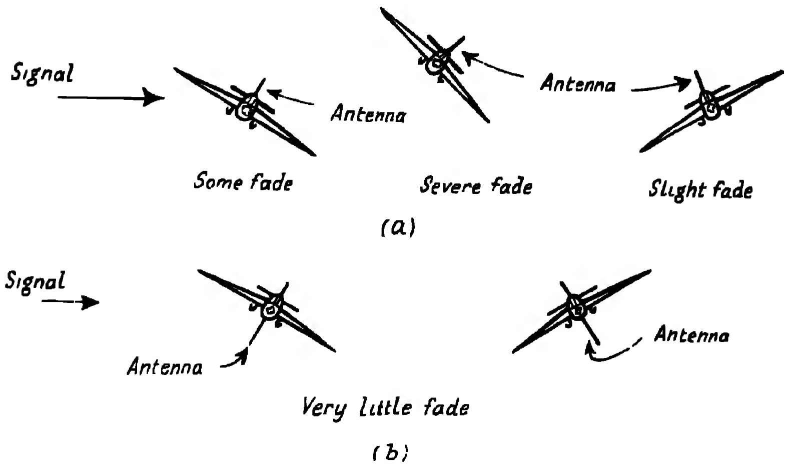



FIG. 261. Effects of aircraft turn and bank on received signal intensity. (a) Antenna mounted on top of fuselage. (b) Antenna mounted below fuselage.

the plane is shown. Almost no change in signal level is noticeable while circling. Refraction is not severe when the wing is between the antenna and the station, because the refraction angle over the trailing edge of the wing would have to be very great with the antenna mounted on the lower side of the wing almost as far forward as the leading edge. Small changes in signal due to smaller projected areas when at right angles to the station occur.

## SKIP DISTANCE

**The Kennelly-Heaviside Layer.** The presence of a reflecting medium in the atmosphere above the surface of the earth has long been recognized by scientists as a phenomenon affecting radio communication. Until the time of Marconi's first transoceanic experiment, it was theorized that radio waves, being electromagnetic in nature, were of the same character as heat and light waves. It was therefore assumed that, in common with such waves, radio waves traveled in straight lines with no more than the customary deviations due to diffraction, reflection, and refraction. When Marconi successfully demonstrated the transmission of radio signals across the Atlantic from Clifden, Ireland, to Glace Bay, Nova Scotia, in

1901–1902, it was evident that this theory was erroneous. If the radio waves traveled in a straight line during this historic experiment, it must be assumed that they penetrated the earth's crust. This possibility was easily disproved mathematically. The alternative possibility, that diffraction caused the radio waves to bend sufficiently to follow the curvature of the earth, was also proved false.

In 1902, the American scientist, Dr. A E. Kennelly, and Oliver Heaviside, a British scientist, suggested the existence of a reflecting medium in the upper layers of the earth's atmosphere to account for this peculiar behavior of radio waves. Each of these men, working independently, arrived at this conclusion mathematically. Their propositions were

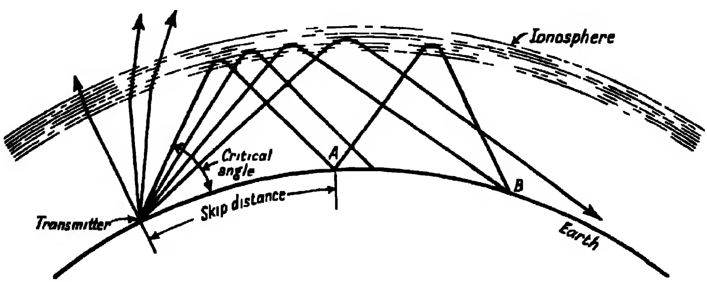



FIG. 262. Effect of the ionosphere on the sky wave component of a radiated signal

announced almost simultaneously. A great number of different investigators analyzed the theoretical possibilities of such a medium, or layer, in the upper atmosphere, and this theory was eventually accepted as a satisfactory explanation for the return of radio waves to the earth. It is interesting to note that the existence of this reflecting layer was predicted by Kennelly and Heaviside many years before long-distance h-f radio communication demonstrated its proof The layer was subsequently named the **Kennelly-Heaviside layer** in honor of the two scientists.

As the height above the earth increases, the air becomes more and more rare, and the intensity of the sun's radiation increases. The lower air pressure and the increased radiation both cause a tendency toward ionization of the atmosphere. The air molecules of the rarefied atmosphere are ionized as a result of bombardment by cosmic and solar radiation, so that free electrons are present in the ionized region. The cloud of electrons forms a conducting medium which causes radio waves to be refracted, or reflected, back to the earth.

This ionized region of the atmosphere constitutes the Kennelly-Heaviside layer and extends from 70 to 250 miles above the earth's surface. Because of its nature, it is also often called the **ionosphere.**

The ionosphere is responsible for the behavior of the sky-wave component of a radiated signal. Since an antenna radiates waves at all

angles, a major portion of the field energy is expended in radiation skyward. Were it not for the ionosphere, all this energy would continue traveling outward into space in a straight line and would be lost. The ionosphere, however, acts as a reflector and causes the radio waves to be bent back to earth. The angle at which this reflection occurs depends upon the angle at which the wave has left the transmitting antenna. The latter is called the **radiation angle,** or **wave angle.** The wave angle is very important, since it determines the general area on the surface of the earth at which the signal will return. This is shown in Fig. 262.

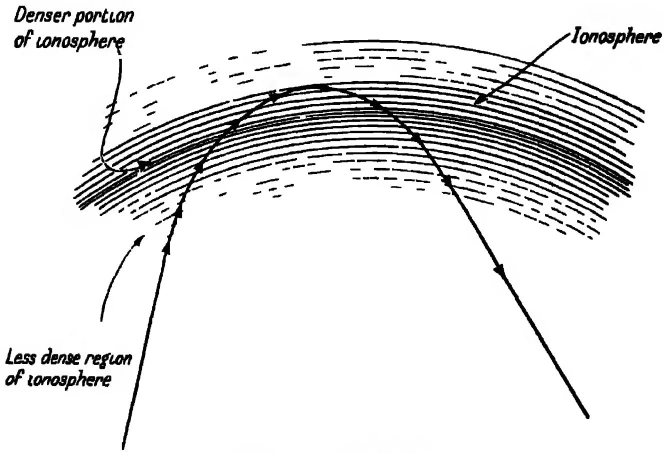



FIG. 263 Ionosphere refraction

Actually, a radio wave is refracted as much as reflected by the ionosphere. At the lower frequencies, the action is mainly reflection. As the frequency is increased, more and more refraction occurs, so that the wave is gradually bent around until it proceeds earthward again. In reflection, the wave strikes the conducting medium of the ionosphere and is reradiated in the same way that light waves are reflected from a mirror. In refraction, the action is also similar to that of light waves when passing through mediums of different density. It is a well-known fact that when light waves pass from a given medium into a denser medium, the waves are bent. This can be observed by placing a pencil in a glass of water. The apparent crookedness of the pencil is due to the refraction of the light waves as they pass through the denser water (traveling from the immersed part of the pencil to the eye) compared with the waves that reach the eye from the upper part of the pencil. In the same manner, as radio waves pass from the atmosphere into the ionosphere, which is denser by virtue of its electron cloud, they are bent. As the waves progress farther into the ionosphere, the electron density increases, resulting in an increase in the angle of bending. Once the wave is turned

around so that it is heading back to earth, it passes into progressively less dense portions of the ionosphere. The bending, therefore, is in the opposite direction, as shown in Fig. 263. The net result is the same as though the wave had been reflected instead of refracted, so it is customary to speak of a wave being reflected from the ionosphere in both cases. This usage will be adhered to throughout this chapter.

As mentioned above, the angle at which a reflected wave returns to earth depends upon the initial wave angle. As the angle at which a radio wave enters the ionosphere is increased, the wave returns to earth nearer the transmitting point. As the angle is still further increased, a point will be reached, called the **critical angle,** at which the wave just manages to be bent back to earth. Waves entering the ionosphere at angles

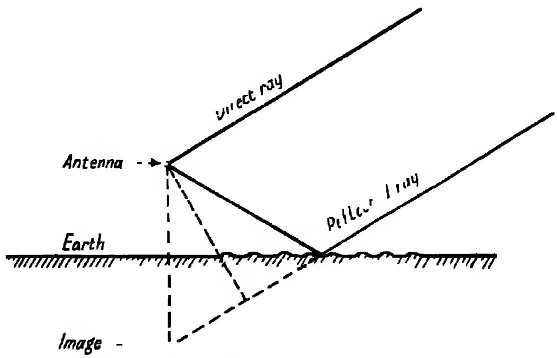



FIG. 264 Effect of ground reflection

beyond the critical angle are not bent enough to return to earth, and as a result they pass entirely through the ionosphere to outer space. Since these waves are obviously useless for radio communication, all waves radiated at angles above the critical angle are wasted and the energy expended in them is a total loss (see Fig. 262)

The critical wave angle varies with frequency, becoming smaller as the frequency is increased. At extremely high frequencies, the bending effect is so small that no waves return to the earth.

It is apparent that in order to secure efficient radio communication over a given distance, maximum radiation should be concentrated at the proper wave angle. The performance of an antenna is considerably modified by the presence of the earth underneath it because of the reflection characteristic of the earth, as discussed in the section on image antennas. Waves radiated from the antenna at angles below the horizontal are reflected upward by the earth, as shown in Fig. 264. Waves radiated from the antenna at angles above the horizontal combine with the reflected ground waves in various ways, depending upon the height of the antenna above the earth. At some vertical angles, the direct and

reflected waves may be exactly in phase and therefore additive, with the result that maximum radiation occurs at these vertical angles. At other vertical angles the two waves may be 180° out of phase, resulting in complete cancellation of radiation, or a *null* in the field pattern at this angle. At all other vertical angles, the resultant radiation field is the vector sum of the direct and reflected fields and will therefore have various intermediate values. Thus, it can be seen that the wave angle at

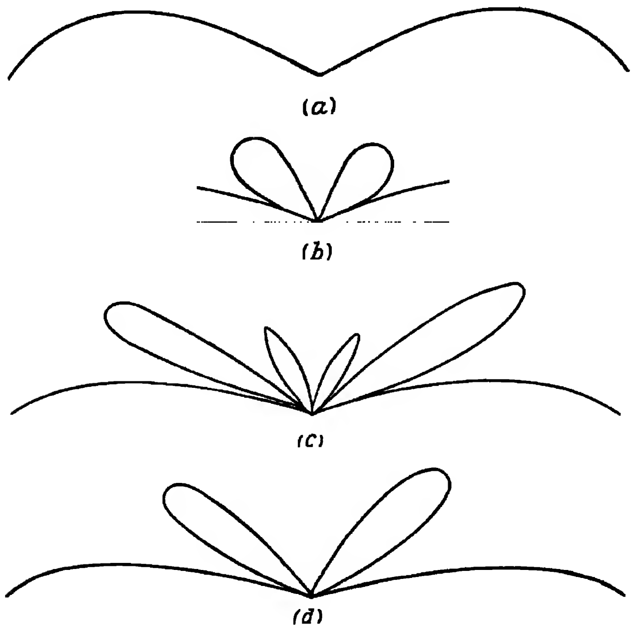



FIG. 265. Vertical-plane radiation patterns for vertical half-wave antenna with centers at different heights above ground. (a) Height ¼ wave length. (b) Height ½ wave length. (c) Height ¾ wave length. (d) Height 1 wave length.

which maximum radiation occurs for a given antenna is a function of the height of the antenna above the earth.

Figure 265 illustrates the vertical-plane radiation patterns for a vertical half-wave antenna at different heights above the ground. The antenna height is that of the center of the antenna and is given in wave lengths. Corresponding data for a horizontal half-wave antenna is shown in Fig. 266 for various heights above the earth. In both figures, perfect earth conductivity is assumed. At all except the lowest wave angles, the effect of ground losses due to imperfect earth conductivity is small, and no serious error is introduced by assuming perfect earth conductivity.

It will be noted that the vertical-plane field pattern varies with the direction of the antenna wire, that is, it depends upon whether the field

is endwise or broadside to the antenna. The variation, however, is in *amplitude* of the field. The effect on optimum wave angle is negligible and, in general, need not be considered as a factor affecting the directivity of directional antenna systems.

**Wave Path.** It is impossible to make calculations of the wave path of a radio signal with any degree of certainty because of lack of knowledge concerning the distribution of the electron density with variations in ionosphere height. A number of hypothetical assumptions have been made, however, based on observed results. The behavior of radio waves of different frequencies has led to the conclusion that there are several ionized layers in the ionosphere rather than only one

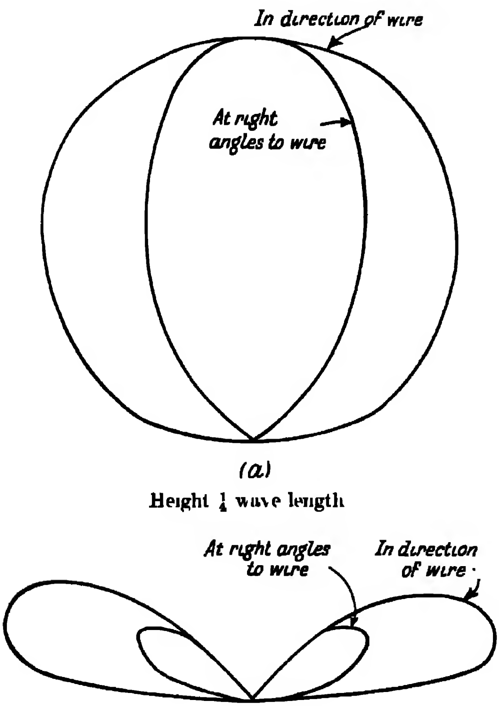
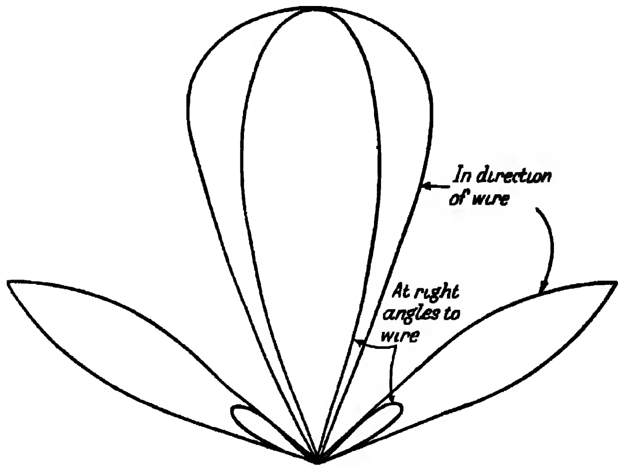



FIG. 266 (*See opposite page for caption.*)

Identifying letters have been assigned to the various layers of the ionosphere, the layer closest to the earth being called the *E* layer. The *E* layer remains at an essentially constant height of 70 miles above the surface of the earth. The next higher layer is the *F* layer. The ionization of both *E* and *F* layers varies in density with the time of day and is at a minimum for three or four hours after midnight. The density of the *E* layer is greatest at noon. The *F* layer splits into two layers about sunrise, the lower layer being called the $F_1$ layer and the upper, the $F_2$ layer. These layers continue as separate layers throughout the day and merge into the single *F* at night. Each of these layers exists at different heights above the earth. The *F* layer height is approximately 185 miles, the $F_1$ 140 miles and the $F_2$ varying from 200 to 250 miles. During the winter months, the $F_1$ layer completely disappears, leaving the single $F_2$ layer during the daytime at an effective height of 140 miles.

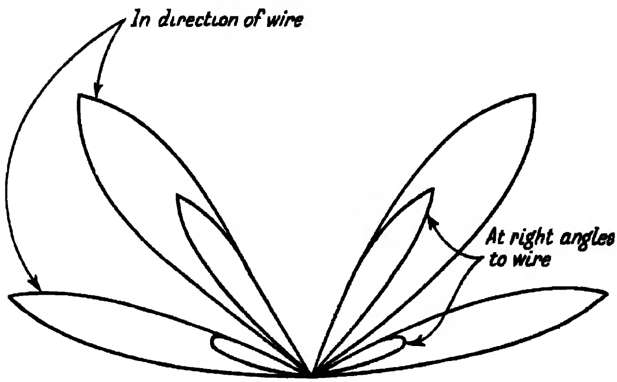
In addition to the variation in layer height and density day by day

as noted above, variations in density of ionization also occur as the sun's radiation varies. There are three such cycles of ionization in addition to the daily variation. One such cycle occurs every 28 days and coincides




(c)

Height $\frac{3}{4}$ wave length.




(d)

Height 1 wave length.

FIG. 266. Vertical plane radiation patterns for horizontal half-wave antennas at different heights above ground.

with the period of the sun's rotation. Another cycle occurs yearly in accordance with the earth's path around the sun. The third cycle has been found to occur every 11 years and has been presumed to coincide with variation in the sun's radiation due to sunspot activity.

The degree of refraction of a radio wave is a function of the ionosphere density. In view of the many changes in ionosphere density discussed above, it is evident that the path of a radio wave depends upon a great many factors.

In addition to the effects of ionosphere density, the degree of refraction is also a function of the frequency of a radio wave. The bending of a wave by the ionosphere has been found to become more acute as the frequency is decreased. This effect is so pronounced that, at very low frequencies, radiation at *all* wave angles, including the absolute vertical (90°), is returned to earth. At frequencies lower than approximately 4 megacycles, the sky wave returns to the earth at a point within the range of the ground wave. On such frequencies, therefore, practically continuous coverage occurs as the distance from the transmitter is increased until the signal is completely attenuated.

At the higher frequencies, the point at which the sky wave returns to earth occurs at increasingly greater distances from the transmitter. Between the outer limit of the ground-wave coverage zone and the beginning of the returned sky-wave zone, an area exists in which no signal is heard. This zone of silence becomes greater as the frequency is increased. The interval between the ground-wave zone and the returned sky-wave zone is known as the **skip distance.**

A radio wave that is reflected back to the earth by the ionosphere is usually reflected again, depending upon the conductivity of the earth at the point of arrival. Thus, a wave may make several hops, or skips, along the surface of the earth before becoming completely attenuated. This *multiple reflection* is caused by the wave being reflected back and forth between the earth and the ionosphere a number of times. Most long-distance communication is accomplished by multiple reflection, since under the most favorable conditions the greatest possible skip distance is about 3,000 miles.

The choice of frequency for communication over a given distance depends upon a number of factors. For example, the highest frequency that will be reflected to the earth within the required distance should be selected. Lower frequencies could be employed, but these would be subject to multiple reflection in traversing the required distance with attendant increase in attenuation. No set rule can be given because of the many variables to be considered. In general, when the distance is very great and the entire path is in daylight, a frequency in the vicinity of 30 megacycles will be satisfactory. For lesser distances, the 30-megacycle skip distance would be too great, and a lower frequency would have to be used. At night, these frequencies would be too high and would probably not be reflected back to the earth at all. A frequency in the vicinity of 10 megacycles would then be indicated When the path to be covered is partly in daylight and partly in darkness, which often occurs

in east to west transmission, a frequency intermediate to the two above frequencies would effect a satisfactory compromise. Uninterrupted communication between two fixed points can usually be assured by the use of three frequencies.

Once an operating frequency has been chosen, the optimum wave angle for this frequency can be chosen by reference to the chart of Fig. 267. This chart gives the maximum wave angle at which a wave of given frequency can be reflected from the ionosphere.

Because of the continuously variable factors affecting the ionosphere, the refraction of several waves from a given transmitter entering the ionosphere at essentially the same angle will not be uniform. A group of such waves may arrive at the receiving station in various conditions of relative phase. At times, they may arrive in phase and add, causing an increase in signal intensity, while at other times, they may tend to cancel each other out, resulting in a decrease in signal intensity, with the signal often disappearing completely. This variation in signal strength may be quite rapid, especially at the higher frequencies, and is called **fading**.

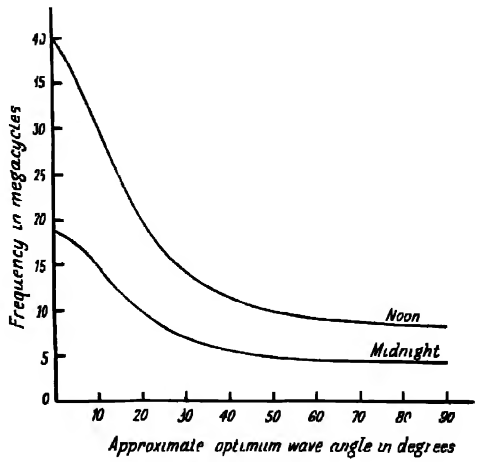


FIG. 267

The effects of fading can be largely eliminated in commercial receiving installations by the use of *diversity receiving antenna systems*. It has been determined empirically that when the same signal is received on separate antennas 10 or more wave lengths apart it will not fade on all antennas simultaneously. When three or more antennas are employed, the possibility of the signal fading out completely on all three antennas is extremely small. Commercial installations with such antenna systems use three separate receivers, one for each antenna, and the combined output of these receivers is fed into a common amplifier system.

## TRANSMISSION LINES

The physical arrangement entailed in practically all radio stations is such that it is usually impossible to connect an antenna directly to the transmitter output circuit. Even where such an arrangement is possible,

it is usually desirable to erect the antenna in the clear, well away from surrounding objects with their attendant losses and detrimental effects upon the antenna effectiveness.

Radio-frequency energy is transmitted from a transmitter to an antenna by means of **transmission lines.** A transmission-line network is known as a **feeder system.** An unfortunate tendency exists in some circles to describe antenna systems according to the feeder system used. Such method of classification is completely erroneous and should be discouraged, since it gives rise to ambiguities. The performance of an antenna is completely independent of the type of feeder system used. For this reason antenna systems and feeder systems are discussed separately in this text.

Transmission lines used in antenna systems are of two general types—*resonant transmission lines* and *nonresonant transmission lines.* Resonant lines are also often called "tuned lines."

**The Resonant Transmission Line.** If a transmission line is infinitely long, power applied to it will eventually be dissipated and there will be no reflection or standing waves on the line. Every line has a *characteristic*, or *surge*, impedance, depending upon the size of the conductors, the spacing between them, the capacitance between the lines per unit length, and the inductance per unit length. If a transmission line is cut to a specific length and the line is terminated in a resistance equal to the characteristic impedance, there will also be no standing waves, because, for the section of line under consideration, nothing has been changed. The terminating resistance absorbs an amount of energy equivalent to the absorption of a line infinite in length, since the terminating resistance is equal to the characteristic impedance of this line if it were infinite in length. It must be remembered that the line should terminate in a pure *resistance*. Any reactive component will result in an impedance which does not match the surge impedance of the line.

If a given transmission line is not terminated in its characteristic impedance, standing waves of voltage and current will appear on the line, and the amplitude of these standing waves will be a function of the degree of mismatch. The ratio of the standing wave maxima, or loops, to the standing-wave minima, or nodes, is equal to the ratio of the characteristic impedance to the actual terminating impedance. Thus, if a line having 500 ohm impedance is terminated in 50 ohms, the *standing-wave ratio* will be 10. Similarly, if the line is terminated in 5,000 ohms, the standing-wave ratio will be 10. In either case, the standing waves on the line will be of the same amplitude, the only difference being in the displacement of the node points along the line.

Since a transmission line of this type is a resonant, or tuned, circuit, power could not be put into the line unless the length were such as to make it resonant at the operating frequency. In most cases, it is

impractical to cut the line to the exact necessary length, so the line is tuned by an inductance-capacitance circuit at the transmitter end, as shown in Fig. 268.

Resonant lines are not used for high-power installations or in installations where the distance is appreciably great because of the losses involved. Although the fields about each of the conductors in a resonant line are 180° out of phase and therefore cancel to a large extent, the losses are considerable if a standing-wave ratio of 10 is exceeded. For low-power equipment where transmission-line lengths are short, such as on aircraft, resonant lines can be used effectively.

**The Nonresonant Transmission Line.** Nonresonant transmission lines are lines which are terminated in their characteristic impedance and therefore have no standing waves on them. They have the major advantage that they may be cut to any length. When a transmission

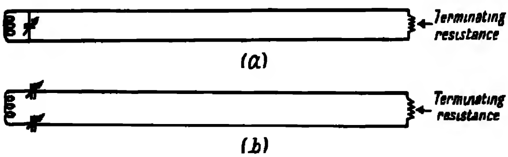



FIG. 268 Methods of tuning transmission lines

line is of the open wire type, that is, when the dielectric is air, the characteristic impedance for the line can be computed from the formula

$$Z = 276 \log \frac{b}{a}. \tag{4}$$

where $Z$ characteristic (or surge) impedance in ohms
$b$ center-to-center spacing of wires;
$a$ radius of the conductor;
276 a constant.

The values of $b$ and $a$ can be in any convenient unit, such as inches or centimeters, but both values must be in terms of the same unit.

Commercial antenna installations make use of concentric, or coaxial, transmission lines. Such lines, although expensive, are very efficient and have low loss. They consist of a wire located in the center of a conducting tube, the wire forming one conductor of the line and the tube the other. Concentric lines are fitted with special end seals, and the air in them is replaced by nitrogen or some other inert gas. This prevents condensation of moisture within the line and the electrical leakage that would result.

The characteristic impedance of a concentric transmission line is determined from the formula

$$Z = 138 \log \frac{b}{a}, \tag{5}$$

where $Z$ = characteristic (or surge) impedance in ohms;
$b$ = inside diameter of outer conductor;
$a$ = outside diameter of inner conductor;
138 = a constant.

This formula assumes air dielectric throughout the line. In actual practice, of course, the inner conductor is supported by ceramic spacers

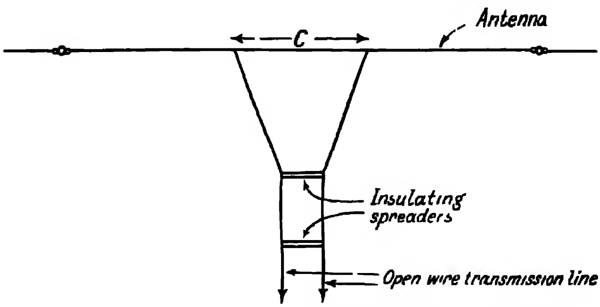



FIG. 269. Delta or Y antenna transmission-line coupling.

placed at intervals along the line. If the spacers are widely spaced, however, the error is negligible.

The proper terminating impedance for a nonresonant line is obtained by connecting the line to the antenna at a point where the impedance equals the characteristic impedance of the line. Low-impedance (55 to 100 ohms) lines can be connected directly to the antenna at a point of current loop. The popular twisted pair lines and concentric lines are of this type.

Open-wire lines usually have impedances on the order of 400 to 700 ohms. Lower impedance open-wire lines are impractical because of the extremely close spacing required. One very popular method of achieving a match at the antenna end is to employ the *delta* coupling shown in Fig. 269. The final section of the transmission line is fanned out until its spacing gives the desired impedance. The antenna end of this line section will then have an impedance equal to the antenna at this point (across $C$, Fig. 269). The lower end of the line section will match the impedance of the transmission line.

Another method of matching transmission-line impedance to that of the antenna is by the use of so-called **quarter-wave stubs,** or **matching sections.** In this arrangement a quarter-wave length of transmission

line acts as a step-up or step-down transformer. It has been shown that any two resistances that do not differ too greatly may be matched by inserting between them a quarter-wave length line having the proper characteristic impedance. The required characteristic impedance of such a line section is obtained from the equation

$$Z = \sqrt{Z_A Z_L}, \tag{6}$$

where $Z$ = surge impedance of matching stub;
$Z_A$ = antenna impedance;
$Z_L$ = surge impedance of transmission line.

It should be remembered that since the antenna is operated at resonance, $Z = R$ and the antenna will present a pure resistance load. Thus, if the antenna resistance is 80 ohms and it is desired to match a 500-ohm nonresonant line to it, a quarter-wave stub having an impedance obtained from Eq. (6) should be used, or

$$Z = \sqrt{500 \cdot 80}, \tag{7}$$

$$Z = 200 \text{ ohms}. \tag{8}$$

If the available wire size is known, the spacing required in the matching section can be obtained from Eq. (4).

## *COUPLING METHODS*

The factors to be considered in transmitter-antenna coupling circuits are *maximum* transfer of energy at the operating frequency and *minimum* transfer of energy at any other frequencies, notably the harmonics. As was discussed in a previous chapter, maximum transfer of power between two circuits occurs when the impedance of the source is equal to the impedance of the load. If the antenna is fed by a nonresonant transmission line properly matched and balanced, the load impedance presented to the final stage of the transmitter will equal the surge impedance of the transmission line. All that is necessary, therefore, is to match the impedance of the final transmitter-stage output circuit to the characteristic impedance of the line, which is most efficiently accomplished by means of a transformer of the proper turns ratio. A number of representative circuits of this type are shown in Fig. 270.

The second harmonic content in any amplifier system is greater in amplitude than any other harmonic component. Most high-power transmitters have push-pull power-amplifier stages. If the final push-pull stage is properly balanced, the second-harmonic content is balanced out. Nevertheless, considerable second harmonic current flows in the final tank inductance, but is prevented from being transferred to the antenna coil by the insertion of a grounded electrostatic shield, as shown in Fig.

270(a). Another method which is effective in the suppression of harmonics is *link coupling*, shown in Fig. 270(b). The wide separation between antenna and transmitter circuits minimizes electrostatic coupling, which is the only path by which harmonics from a push-pull amplifier can be transferred.

The circuit shown at (c) in Fig. 270 makes use of series resonant circuits to by-pass undesirable harmonic currents to ground. The resonant

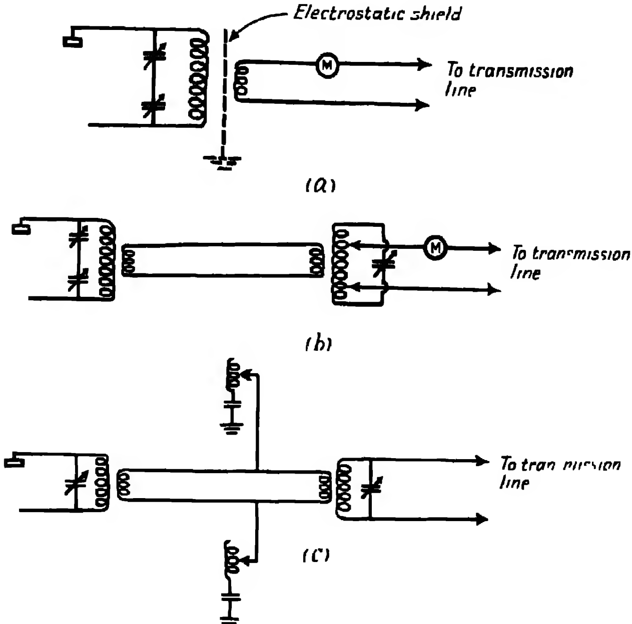



FIG. 270. Antenna coupling systems. (a) Reduction of second harmonic output by electrostatic shield. (b) Link coupling. (c) Link coupling arrangement incorporating harmonic reducing network.

circuits are tuned to the second harmonic and offer a low-impedance path to ground at this frequency while the fundamental is comparatively unaffected. The circuit at (c) in Fig. 270 utilizes a series resonant circuit in *series* with the antenna coupling circuit. This resonant circuit is tuned to the fundamental frequency and offers low impedance to it. At all other frequencies, including the harmonics, the impedance is high.

It is required by international law that all radio stations reduce their harmonic output as much as "the state of the art" permits. This is a very broad statement and does not allow for the reduction in efficiency that 100 per cent harmonic suppression would effect. In general, a radiated wave may be considered free of harmonics if the second harmonic content does not exceed 0.02 per cent of the carrier at a distance of

approximately 1 mile from the transmitter. In high-power transmitters, a second harmonic content of 0.05 per cent is acceptable.

## *QUESTIONS AND PROBLEMS**

**1.** What is the directional reception pattern of a vertical antenna?

**2.** What is the primary reason for terminating a transmission line in an impedance equal to the characteristic impedance of the line?

**3.** In general, what type of antenna is most suitable for broadcast stations?

**4.** If a vertical antenna has a resistance of 500 ohms and a reactance of zero at its base and antenna power input of 10 kw, what is the peak voltage to ground under 100 per cent modulation?

**5.** If a vertical antenna is 405 ft high and is operated at 1,250 kc, what is its physical height expressed in wave lengths (1 m equals 3.28 ft)?

**6.** What must be the height of a vertical radiator $\frac{1}{2}$ wave length high if the operating frequency is 1,100 kc?

**7.** What is a dummy antenna?

**8.** What is the antenna current when a transmitter is delivering 900 w into an antenna having a resistance of 16 ohms?

**9.** If the day input power to a certain broadcast-station antenna having a resistance of 20 ohms is 200 w, what would be the night input power if the antenna current were cut in half?

**10.** A long transmission line delivers 10 kw into an antenna; at the transmitter end the line current is 5 amp, and at the coupling house it is 4.8 amp. Assuming the line to be properly terminated and the losses in the coupling system negligible, what is the power lost in the line?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

# Chapter XVI

# RADIO AIDS TO NAVIGATION

The directional characteristics of antennas have already been discussed in Chap. XV. Such characteristics are fixed for a given transmitting or receiving antenna; that is, once the antenna is installed, the direction of maximum sensitivity or radiation cannot be changed. The development of the *loop antenna* made possible the variation of directional characteristics at will by actual physical displacement of the antenna itself. By using a sensitive receiver in connection with a loop antenna, it

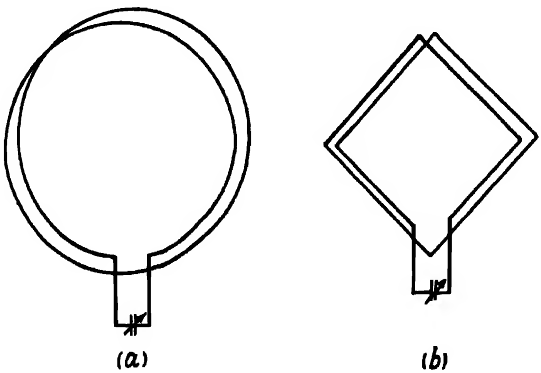



FIG. 271

is possible to ascertain the direction from which a radio signal is arriving with relatively good efficiency. Such a system is called a **radio direction finder** and finds wide application among marine and air navigators as a means of obtaining bearings from a known point when the use of visual sights is impossible because of weather conditions or too great distance.

## *THE RADIO DIRECTION FINDER*

**Principle of Loop Antennas.** The heart of a direction-finding system is the loop antenna. A **loop antenna** is essentially a large coil, and, although it may be of any convenient shape, the most commonly used loops are circular or square in section, as shown in Fig. 271. In conjunction with the tuning capacitor across its input terminals, the loop forms a tuned resonant circuit. The number of turns in the loop coil, therefore, depends

upon the frequencies that are to be received and the frequency range to be covered.

The theory of the loop is best described by considering the case of a square, or rectangular, loop antenna in the path of a radio wave. When the plane of the loop is perpendicular to the direction of wave travel, that is, broadside to the oncoming wave, the voltages induced in the two legs are in phase and of equal magnitude, as shown in Fig. 272(a). The resulting current flow will be in opposite directions around the loop. Since the currents are equal, they will cancel each other, resulting in zero response.

When the plane of the loop is parallel to the direction of wave travel,

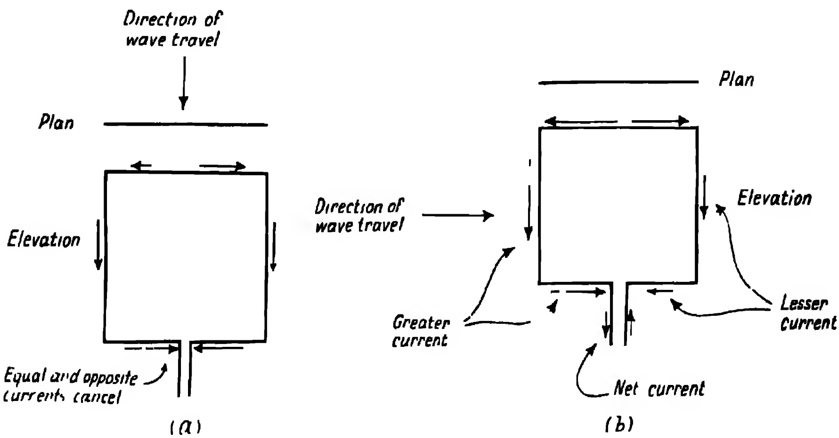



FIG. 272

the wave front reaches the two legs at slightly different times. The voltages induced in the two legs are therefore unequal at any instant and differ slightly in phase. The loop voltage, then, is the vector sum of these two voltages. As the loop is rotated toward the previous perpendicular position, this resultant voltage becomes smaller until it reaches zero at the perpendicular position; but when the plane of the loop is parallel to the direction of wave travel, the resultant voltage is greatest. Maximum response, or sensitivity, therefore results at the parallel position, as shown in Fig. 272(b).

The sensitivity of a loop antenna varies directly with the cosine of the angle that the loop plane makes with the direction of wave travel, as shown in Fig. 273. The plane of the loop is depicted as a single line as viewed from directly above. The vertical legs appear as the points $A$ and $B$, and point $O$ represents the center, or axis, about which the loop rotates. As has been shown, the response of a loop depends upon the magnitude and phase of the voltages developed in the two legs by the

radio wave. The latter, in turn, depends upon the *effective distance* of the legs from the loop center *in the direction of wave travel.* Since the loop is symmetrical, the distance of either leg from the center is equal to the distance of the remaining leg from the center. The loop response can therefore be conveniently taken as a direct function of the effective distance of one of the loop legs from the loop center.

In Fig. 273(a). the loop is at maximal position. The effective distance of leg $B$ from the loop center in the direction of wave travel is therefore equal to the actual distance, or $OB$. In this position, the plane of the loop is parallel to the direction of wave travel, and the loop consequently makes an angle of 0° with the wave direction. Since the cosine of 0° is 1 (Appendix Table III), maximum loop response is indicated by unity.

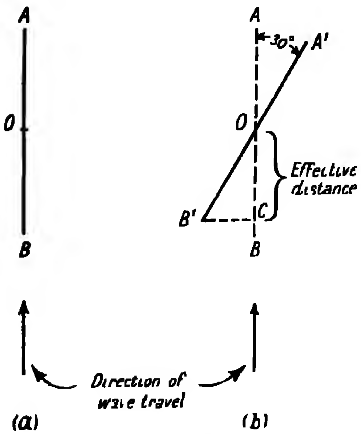


FIG. 273.

When the loop is rotated through 30°, as shown in Fig. 273(b), the voltage induced in leg $B'$ with respect to loop center $O$ will be of the same phase and magnitude as if the loop had been contracted in size to the dimension $OC'$, since only a change in distance *along the line of wave travel* will effect a change in magnitude and phase of the induced loop voltage. The distance $OC'$ can therefore be taken as the *effective distance* of leg $B'$ from the loop center at the 30° position. The sensitivity of the loop was at maximum at position $OB$. The sensitivity at the 30° position is in the same proportion to maximum sensitivity as the length of line $OC'$ is to the length of line $OB$. Expressed mathematically,

$$\text{sensitivity at } 30° = \frac{OC'}{OB}. \tag{1}$$

Since the distances $OB$ and $OB'$ are identical, Eq. (1) becomes

$$\text{sensitivity at } 30° = \frac{OC'}{OB'}. \tag{2}$$

However, in triangle $OB'C'$, angle $B'OC$ is 30°, and by trigonometry

$$\frac{OC'}{OB'} = \cos 30°. \tag{3}$$

Substituting in Eq. (2),

$$\text{sensitivity at } 30° = \cos 30° = 0.88603. \tag{4}$$

Equation (4) gives the sensitivity at the 30° position as a proportion of maximum sensitivity. Since the sensitivity at maximum position has been taken as unity (cosine of 0° equals 1), the sensitivity at the 30° position is the 0.88603 part of 1. Stated in another way, the sensitivity at 30° is 86.603 per cent of maximum sensitivity.

Loop sensitivities at 0°, 30°, 45°, 60°, and 90° are shown in Fig. 274. In Fig. 274(e), the loop is in the 90° position, that is, perpendicular to

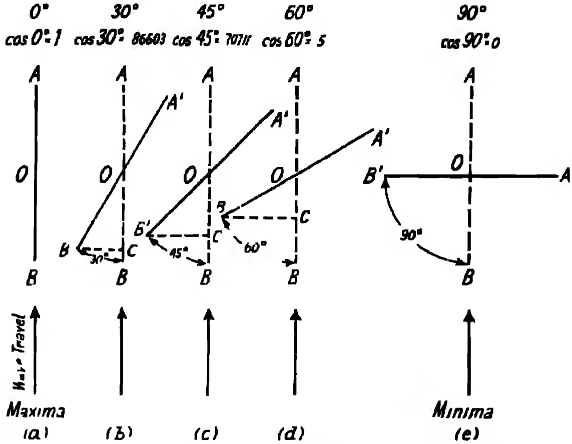



FIG. 274. Variation of loop sensitivity with rotation.

the direction of wave travel. Consequently, the distance $OC'$ becomes zero, and cosine 90° becomes

$$\cos 90° = \frac{0}{OB'}. \tag{5}$$

Loop sensitivity at 90° is therefore zero, or minimum.

The variation of loop sensitivity with the degree of loop rotation can be seen in Fig. 274. Thus, for a rotation of 30° from zero position, the loop sensitivity decreases from the maximum of 1 to 0.88603. This represents a decrease in signal strength of approximately 13 per cent. For a similar amount of rotation from minimum position (rotating from 90° to 60°), the loop sensitivity increases from 0 to 0.5, which represents an increase of 50 per cent in signal strength. For the same amount of loop movement (30°), the loop sensitivity changed 50 per cent from the minimum position compared with only 13 per cent from the maximum position. It is evident, therefore, that it is possible to obtain a much sharper minimum than maximum with the loop antenna.

The variation in sensitivity for any degree of loop rotation can be taken directly from a table of natural cosines. Thus, rotating the loop through 5° from maximum position (rotating from 0 to 5°) decreases the

sensitivity from 1.0 to 0.99619, a change of only 0.3 per cent. Rotating the loop through 5° from minimum position (rotating from 90° to 85°) increases the sensitivity from 0 to 0.08716, a change of 8.7 per cent. For this reason the minimal position is used in taking radio bearings.

There are a number of factors that may cause error in bearings obtained with the radio direction finder. These include unbalance (electrostatically) to ground, presence of conducting objects in the loop vicinity, and reception of horizontally polarized sky waves. Since the last factor is especially troublesome at night, it is often referred to as **night effect.** These factors cause spurious bearings due to shifting of the null, or minimal point, from its proper position and often make it impossible to obtain any minimal position at all.

**Loop Balance.** When a loop antenna is electrostatically unbalanced to ground, the voltages in the two vertical legs will not be equal with

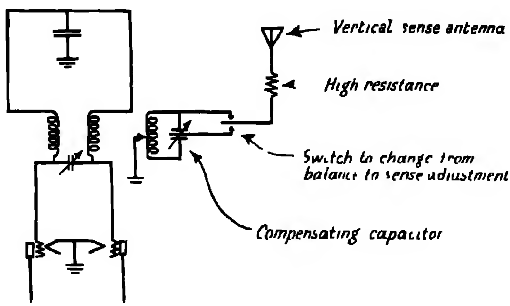



FIG. 275

respect to ground and therefore will not cancel out, because one side of the loop has a larger capacitance to ground with a resultant larger current flowing through this capacitance. The larger current causes a signal to be delivered to the receiver even though the loop is in the zero-response position. In some cases, loop balance may be obtained at other than the true zero position.

Errors from loop unbalance can be minimized by utilizing receiver circuits that are symmetrical with respect to ground. For this reason, most direction-finder receivers use balanced push pull input r-f stages. Typical input circuits are shown in Figures 275, 276, and 277. In the circuit of Fig. 275, residual unbalance in the loop existing despite the symmetrical input circuit is compensated by introducing a voltage from an additional vertical antenna. By means of a three-section condenser (two stators and one rotor), the additional voltage may be coupled to either side of the loop circuit, permitting accurate adjustment of the loop null point. Correct setting is indicated by the adjustment at which a loop rotation of 180° does not affect the null. The vertical-antenna switch is utilized to switch this antenna from balance connection to

*sense-antenna* connection. This function is discussed in a following section.

In addition to the symmetrical circuits and connections discussed in the foregoing paragraphs, the use of an electrostatic shield about the entire loop antenna will further decrease loop unbalance. Such a shield ensures the same capacity to ground for all parts of the loop, regardless of the position of the loop. An insulated bushing is inserted in this metal loop housing (see Fig. 277) to prevent the shield from acting as a single short-circuited turn. The shielded loop is sensitive only to the magnetic component of the radiation field.

**Loop Calibration.** Wires and other conducting objects in the vicinity of a loop, such as metal stays and cargo booms aboard ships, and wings, struts, and so on, aboard aircraft, extract energy from radio waves. As

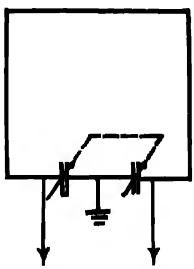


FIG. 276

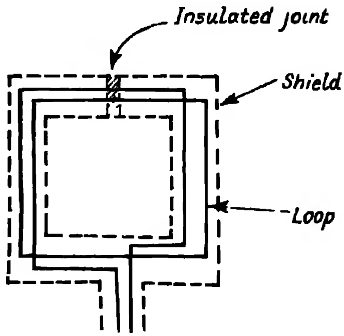


FIG. 277

a result, induction fields are produced by these objects and induce spurious voltages in the loop circuit. Loop minimal positions consequently vary from the proper setting when obtaining bearings from certain directions. For this reason, each radio direction finder installation must be separately calibrated aboard the ship or aircraft upon which it is installed. The calibration must be made with the identical conditions under which the direction finder will normally be used. Thus, aboard an airplane, the calibration should be made with the ship in flight, landing gear retracted and all other radio antenna and external gear secured or in normal position.

Calibration is accomplished by making simultaneous sight and radio bearings at intervals of a few degrees throughout 360° of rotation of the ship or aircraft with respect to a fixed transmitting station. Aboard ship, simultaneous pelorus and radio bearings are usually taken of a lightship while the steamer swings through 360 compass degrees. The procedure is similar with aircraft, the airplane maneuvering through 360 compass degrees while simultaneous sight and radio bearings are

taken of a ground radio transmitting station. Pelorus bearings in this case are taken of the transmitting-station radio tower.

From the foregoing sets of bearings, the radio bearing error as checked against sight bearings is obtained for a number of points throughout 360°. A correction curve can therefore be plotted giving the continuous error of the loop throughout the 360°.

Radio bearings are customarily taken from a loop pointer mounted over a compass card in a conventional compass binnacle. The more modern ships utilize a gyrocompass instead of the usual magnetic compass. Because radio bearings are taken at loop minimum, the loop pointer is normally set at 90° from the plane of the loop antenna. If no external errors existed, this 90° relationship between loop and pointer would be maintained throughout the entire circle of rotation. Practical radio direction finders, however, permit the adjustment of the pointer with respect to the loop at various positions of the circle by means of a cam arrangement. Once the correction curve has been obtained as outlined above, the various cams throughout the circle of rotation are set to correspond with the necessary correction as determined from the curve. The relation of pointer to loop will then vary from 90° by the proper amount throughout the circle and thus compensate for any error introduced by fixed objects aboard ship.

**Night Effect.** Horizontally polarized waves traveling in a downward direction are often present and result in the introduction of induced voltages in the horizontal portions of a loop antenna. Such voltages prevent complete cancellation of the loop voltages at the true null position. Hence, zero resultant voltage is not obtained when the plane of the loop is perpendicular to the direction of wave travel, and as a result of the additional horizontal voltage components, minimum signal often occurs at false positions and usually varies continuously. Since downcoming radio waves are produced by ionosphere action (see Chap. XV), this effect is particularly noticeable at night, and for this reason it is often called **night effect.** Night effect determines the ultimate usefulness of the type of loop antenna under discussion as a direction finder. To date, no method has been discovered to overcome this disadvantage in the conventional loop antenna. A special type of loop antenna, called the **Adcock antenna,** is not subject to the effects of downcoming horizontally polarized waves. In its simplest form, the Adcock antenna consists of two vertical antennas so coupled that there is no effective horizontal component. Because this antenna is, in effect, a single-turn loop, the size of antenna required for sufficient energy pickup prohibits its use in most applications.

**The Sense Antenna.** The loop antenna is fundamentally a bidirectional device. Thus, for a given radio wave, there are two positions of the loop 180° apart at which a null indication can be obtained, since the loop

plane is perpendicular to the direction of wave travel at either of these positions. Usually, this bidirectional characteristic is unimportant, because the transmitting-station bearing is customarily known within 180°. Thus, if an east-west bearing is obtained aboard a ship at sea, the true bearing is easily ascertained because the general direction of the shore from the ship is usually known. Nevertheless, there are a number of situations where an absolute unidirectional bearing is desirable, especially aboard aircraft.

A radio direction finder can be made a true unidirectional device by the use of a small nondirectional vertical antenna. When used for this purpose, such an antenna is called a "sense antenna." As previously discussed, the loop response at maximum signal position depends upon the vector difference between the voltages induced in the vertical legs. If the signal voltage picked up by the vertical antenna is added to the leg of the loop that is *toward* the transmitting station at maximum position, the effective loop response is increased. If the vertical antenna is coupled to the side of the loop *away* from the transmitting station, the effective loop response is decreased.

The sense antenna is coupled to the loop through a high series resistance, which has the effect of greatly decreasing the sense-antenna sensitivity. Were it not for this resistance, the voltage added to *either* loop leg by the vertical antenna would be so great that an increase of signal would result in either of the loop maximum positions. When the proper amount of resistance is used, just enough voltage is added to the loop leg *away* from the transmitting station to balance out the normally greater voltage in the leg *toward* the transmitting station.

In normal use without sense antenna, the loop response depends upon the amount by which the voltage in the *forward* vertical leg (the leg *toward* the transmitter) exceeds (vectorially) the voltage in the remaining vertical leg. Coupling the sense antenna to the forward vertical leg of the loop increases the voltage in this leg. Since the voltage in this leg is already greater than that in the other leg, the result is an increase in signal strength, since the vector difference between the two leg voltages has been increased. This increase occurs despite the fact that the additional voltage introduced by the sense antenna is comparatively small owing to the resistance in the circuit.

In obtaining a unidirectional bearing, the procedure is as follows: The bidirectional bearing is obtained by rotation to the null point in the conventional manner. The loop is then rotated 90° to maximum position *in a certain direction*. The direction in which to turn varies with the particular type of instrument and is specified by the manufacturer. The sense antenna is then coupled to the loop by a switch. If the signal increases, the signal direction, or "sense," as it is called, is in one direction. If the signal decreases, the sense is in the opposite direction. The null

position, which, of the two, is the correct bearing when an increase in signal results from the foregoing procedure, is specified by the manufacturer, since it depends upon which of the vertical legs the sense antenna is coupled to when the switch is closed.

The United States government has established a chain of radio stations along both the Atlantic and Pacific coasts to facilitate the procurement of radio bearings by the use of radio direction finders. These stations, although they serve as beacons to the mariner, are not radio beacon stations in the technical sense. The signals are broadcast in all directions by the stations, and the transmitting antennas have no directional characteristic. Charts showing the locations and schedules of these stations are distributed through the U.S. Coast Guard. Such stations transmit an identifying signal at regular intervals during a certain period each hour of the day. The identifying signal usually consists of a letter of the Morse code and is transmitted very slowly so that persons unfamiliar with the code can easily recognize them. Stations located in the same general area have their schedules staggered throughout each hour in order to avoid possible conflict

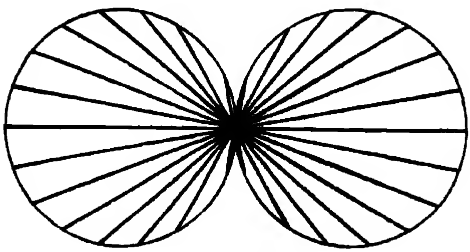


Fig. 278.

## THE AIRWAY RADIO BEACON

The field-intensity pattern of a loop antenna is essentially two tangent circles, as shown in Fig. 278. It will be seen that radiation from a transmitting loop antenna is suppressed in both directions at right angles to the plane of the loop while maximum radiation is obtained in the directions in line with the plane of the loop. If two such loop antennas were installed at right angles to each other, the resulting field pattern would appear as shown in Fig 279. In the areas in which the adjacent loop patterns overlap, the signals from both loop antennas are audible.

Airway beacon, or *radio-range*, stations utilize such an antenna system to assist in the navigation of aircraft by radio. A radio transmitter feeds energy modulated at audio frequencies first to one loop and then to the other by an automatic motor-driven switch. This energy is keyed to produce the Morse code character *N* in the field formed by the *N* antenna in Fig. 279 and the character *A* in the field of the *A* antenna.

The relative disposition of the keyed characters is shown in Fig. 280. The energy above the center line is directed to the *N* antenna, and the

energy below the line is directed to the $A$ antenna. It will be seen that as the energy is switched from loop to loop for the intervals shown, the

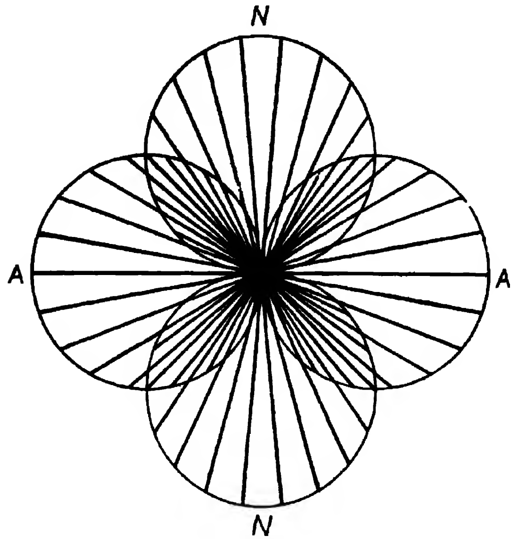



FIG. 279

letters $N$ and $A$ will be formed by the field patterns of the respective loops.

An airplane flying within an $N$ field, or *N-quadrant*, will receive a constant repetition of the letter $N$. If the plane is within an $A$ quadrant,

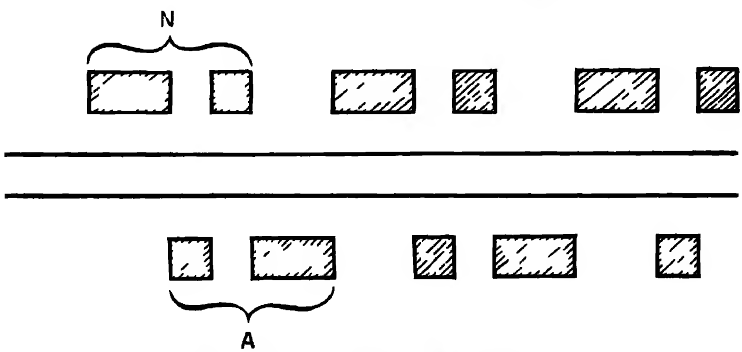



FIG. 280. $A$ and $N$ signal alternation

the letter $A$ will be heard continuously. It follows that if the plane enters the shaded area within two intersecting quadrants, both $N$ and $A$ will be heard, but one letter will predominate, depending upon which quadrant the airplane is in. If the airplane is in the central portion of the shaded area, the signals $N$ and $A$ will be received with equal strength. The dots

and dashes forming one letter will merge with the spaces of the other, and a continuous note will be heard, because the dash of the *N* is transmitted first, then the dot of the *A*, then the dot of the *N*, and then the dash of the *A*. The same would be true at any other point on a line drawn through the center of each of the four zones of overlapping signals. Because of the inefficiency of the human ear in detecting small changes in signal intensity, there is a considerable area on each side of the line of equal signal strength in which the signal is actually heard as a continuous tone. The width of this area, or *course*, has been found in practice to average 3° for the average human ear.

The course signs (*A* and *N*) are broadcast for 30 sec, then they are interrupted, and the station identification signals are broadcast twice, once from each loop. The station identification signals require 7 sec to transmit, resulting in a complete sequence every 37 sec. Present systems transmit 12 *A*'s and 12 *N*'s between station identification signals.

Immediately above the range station is an inverted cone-shaped area where the loop signals cancel each other and no signal is radiated. This area is known as the "cone of silence." When a pilot is flying toward the station along the course, the strength of signals gradually becomes stronger until relatively close to the range, where they rapidly increase in strength. As the airplane passes over the range, the signal suddenly disappears as the cone of silence is entered and suddenly reappears as the signal zone is entered on the other side

In radio-range flying, the pilot follows any one of the four courses, commonly called the **beam,** on which the continuous tone is heard. If he hears only the *A* or the *N* transmission, he can tell that he is either to right or to left of the beam. If the *A* or *N* signal strength is increasing, he is flying *toward* the beacon; if decreasing, he is flying *away* from the beacon. In either case, the beam can be picked up by altering the course 90° and maintaining the new course until the beam zone is intercepted. The course is then again altered 90° in the reverse direction, and the aircraft will be flying "on the beam," as it is called, and traveling toward the beacon, or range, station. The direction of travel is further checked by noting the increase in beam-signal strength as the station is approached.

The pilot knows he is passing over the station when the cone of silence is reached. Thereafter, the beam signal decreases in intensity. By successively increasing the receiver volume control, the pilot retains as long as possible the beam of the station that he is leaving and then picks up the beam of the next station on his course.

As installed on the airways, the range assembly is oriented so that the courses are aligned in the directions desired. By correctly distributing the stations geographically and aligning successive courses, a system of airways has been marked out in much the same manner as the development of a modern highway network.

Like the broadcast band, the aeronautical radio range band (200 to 400 kc) must accommodate a great number of stations, and some interference is unavoidable. To eliminate all danger of mistaken identity, each station is assigned an individual one- or two-letter station identification signal. As previously described, these identification signals are transmitted once from each loop between each series of *A* and *N* sequences. Each radio-range station operates on a designated frequency. Therefore, only one radio range can be received at any one place on the

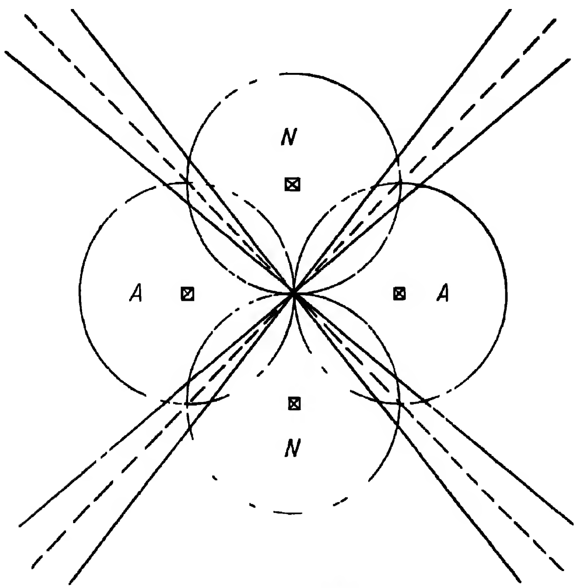



FIG. 281. Horizontal plane field pattern of a radio range station utilizing four vertical antennas

dial. The pilot is consequently provided with two means of identifying a radio range station, by the station frequency or dial setting and by the station identification signals.

**Radio-range Antenna Systems.** The loop type of range described in the foregoing was the earliest practical development of radio-range equipment and was entirely satisfactory for daytime flight operations. With the advent of night flying, the *night effect* previously described in loop antennas precluded satisfactory definition. Modern range stations utilize a system of four vertical antennas in place of two loops. Since the field intensity of a vertical radiator is essentially a circle, the resulting field pattern of four such radiators is equivalent to that obtained with two loops and is shown in Fig. 281.

The loop type of station is relatively inexpensive in initial cost and in

maintenance. For that reason, a great number of loop stations are operated at points where distances in excess of 30 miles are not required. These loop-type stations are useful to fill in gaps along the airway, mark airway intersections, and provide range courses to emergency fields and as localizers for low-approach procedure at airport terminals. With the loop type of station, it is possible to transmit both range signals and weather-broadcast reports, but not simultaneously. The utilization of vertical-antenna systems for range stations permitted the development of simultaneous range stations.

**Simultaneous Range Stations.** In order to transmit a weather report in the loop system, the range signals must be discontinued and the entire

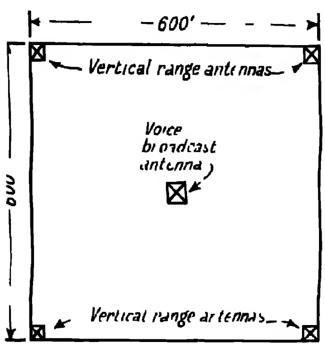



FIG. 282. Disposition of towers in simultaneous range station

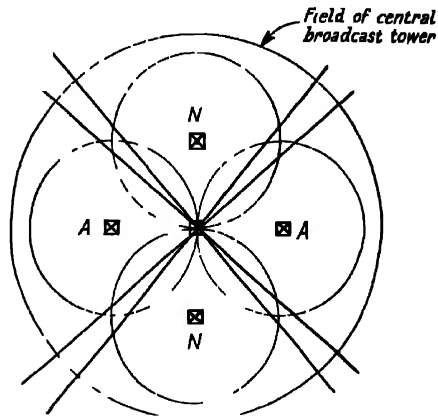



FIG. 283 Field pattern of simultaneous-range station.

system operated as a radiotelephone station. When this is done for the benefit of one group of aircraft, another group may be flying on instruments and depending upon uninterrupted range for accurate navigation. The great number of requests for continuous range operation by pilots on instruments prompted the development of the *simultaneous range station.* Such stations are capable of simultaneously broadcasting range signals and radiotelephone messages on the same frequency.

A typical simultaneous-range station system employs four vertical towers 120 ft high which are placed at each corner of a square measuring 600 ft on a side. Each pair of diagonally opposite towers produces the same figure-8 field-pattern configuration as obtained from a loop antenna, and each such pair is excited by the same keying sequence ($A$ or $N$). In addition to the four corner towers, there is a center tower which operates as a conventional antenna. Figure 282 shows the disposition of the four range towers about the central broadcast tower. The field pattern radiated

by the system is shown in Fig. 283, the largest circle representing the field of the central broadcast tower.

In operation, the center tower continuously radiates an r-f carrier signal at the frequency assigned to the station. The *N* and *A* range signals are radiated from the two pairs of corner towers. These signals are unmodulated and are radio frequencies 1,020 c higher than that of the center tower. A receiver tuned to the assigned frequency of the station will receive signals radiated from the center tower and, in addition, the signals radiated from the four corner towers. The difference of 1,020 c separating these two transmitted signals will cause a beat note of 1,020 c to be developed in the receiver. Since neither the center tower nor the outer signal towers are modulated, no audible response will be obtained

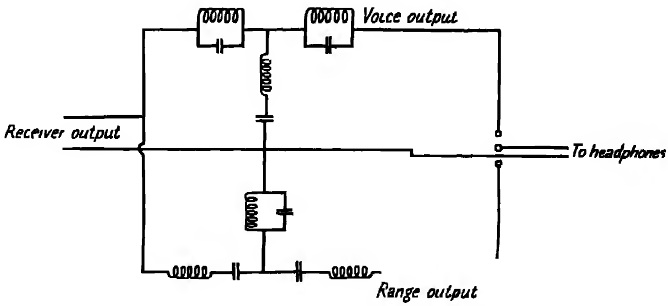



FIG. 284 Simultaneous range receiver output filter circuit

from one or the other of these signals However, the two signals heard together develop an audible beat note corresponding to the frequency separation

The signal resulting from the beat note sounds exactly as though a single frequency transmitter were modulated at 1,020 c. In practical operation it is almost impossible to distinguish between the range signals received from a simultaneous station and those received from a loop-type station.

When radiotelephone messages are to be broadcast, the carrier radiated from the center tower is modulated at regular voice frequencies. The beat-frequency action between the carrier of this center tower and the carriers of the corner towers remains unchanged. In addition to the beat note, the modulation resulting from the voice broadcast is also heard in the earphones by the pilot. In order to eliminate confusion, the radiotelephone and radio range signals are filtered out between the audio output of the receiver and the earphones. The filter is composed of two sections, one to filter out the range signal and the other to filter out the voice signal. Either filter may be switched into the circuit at will.

A typical filter circuit of this type is shown in Fig. 284. Both sections

of the filter consist of inductance-capacitance networks. The range section of the filter network consists of a band-pass filter designed for *minimum* attenuation at 1,020 c with constantly increasing attenuation to approximately 900 and 1,100 c, where sufficient attenuation is obtained to eliminate entirely frequencies below and above these limits. When the earphones are connected to the terminals of this filter section, only the radio-range signals are heard.

The radiotelephone section of the filter consists of a band-elimination filter designed for *maximum* attenuation at 1,020 c with cutoff frequencies of approximately 830 and 1,250 c. The elimination of the 1,020-c component from the voice broadcast causes a small but negligible loss in quality of the speech. When the earphones are connected to the terminals of this filter section, only the radiotelephone signals are heard.

A switching arrangement in the filter network output permits either pilot or co-pilot to listen in any one of three positions—range signal, radiotelephone signal, or both. Under normal conditions, the switch is left in the position where both types of signals can be heard. When no weather reports are being broadcast, the pilot "rides the beam" in this position of the switch. When a weather broadcast commences, the co-pilot can shift his switch to the radiotelephone position and copy the telephone message without interference. At the same time, the pilot can turn his switch to the range position and continue to ride the beam without interference from the voice broadcast.

**Bent Courses.** Radio range courses, which theoretically should be perfectly straight, may be found to have kinks or bands. This condition is most likely to occur where the courses pass close to, or over, hilly or mountainous terrain, large bodies of water, or mineral deposits. Under such conditions a course is sometimes broken up into several parallel courses, usually referred to as **multiple courses.** Multiple courses are extremely difficult to follow because they are very narrow and usually very erratic. They are also extremely confusing to the average pilot who has had limited experience with them. A multiple course may have the same signals on both sides, or it may have the normal signals on either side, or it may have the signals reversed. Bent courses, sometimes called **dog-leg courses,** usually are of little consequence, since the bend generally is small and away from and around the obstruction that caused it. In mountainous country, however, bends have frequently been found that necessitated a change of compass heading of 45° for a short distance in order to stay on course. Several such bends may occur on a given range within a short distance. Obviously, such a range would be hazardous to a pilot who was not familiar with that particular range and its peculiarities. These conditions may be found anywhere, but generally they are confined to hilly or mountainous country. A bent course creates the impression that the course is swinging if the airplane proceeds on a

straight line. Theoretically, the only time that courses actually do swing from their fixed position is usually for a short period at sunrise and sunset. This phenomenon is due to night effect with loop-type range stations and has been practically eliminated with the more modern simultaneous-type range station utilizing vertical radiators.

In many cases, range courses are intentionally shifted to produce a desired beam alignment. Generally, there are two methods of displacing courses from their normal 90° separation. In both methods, shifting is accomplished by a reduction of radiated energy in one or more of the vertical antennas. In the first method, called **course squeezing,** the energy radiated from two diagonally opposite towers is reduced. The resulting field pattern is shown in Fig. 285. Thus, either the fields of both *N* towers or the fields of both *A* towers are reduced, depending upon the result desired. In the illustration, the *N* fields have been reduced, as shown by the heavy dotted lines. The light dotted lines show the original 90° course separation with fields of equal strength. It will be noted that any displacement of course is achieved at the expense of a corresponding reduction of the distance over which the displaced courses may be used.

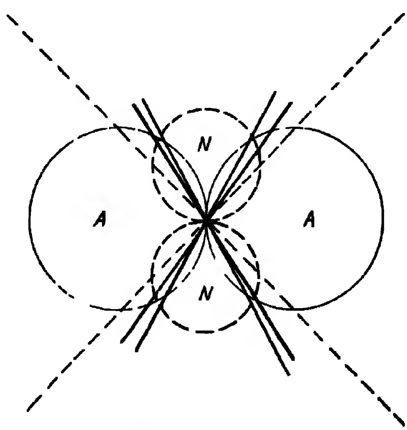


FIG. 285 Course squeezing.

After shifting courses in the manner described above, opposite or reciprocal courses remain 180° apart, that is, they form a straight, or linear, course. The second method of course shifting is called "course bending" and is used when it is desired to create a bend in reciprocal courses. **Course bending** is accomplished by increasing the relative energy radiated from one tower above that of the remaining towers. The energy radiated from the diagonally opposite tower is reduced below that of the remaining towers. This arrangement is shown in Fig. 286. The result is the squeezing of two courses in one quadrant and an increase in separation of the two courses in the opposite quadrant.

**Radio-marker Stations.** Radio markers are designed to indicate to the pilot flying a radio-range course his position along such a course and so to enable him to orient himself with relation to the terrain over which he is flying. Such radio markers are of three types— the M, FM (*not* frequency modulation), and Z.

The M type of marker comprises a low-power radio transmitter with a nondirectional antenna system. All such markers operate on the same

frequency as the range station on whose course they are located. Their signals are received by the pilot as an interference on the range signal without the necessity of retuning his range receiver. Signals from these markers are received as a continuously repeated single letter of the alphabet in Morse code, which identifies the marker, and the power of the markers is such that they may be heard only in the immediate vicinity of the marker. They provide only a general check on position, since the distance over which they may be heard is affected by differences in antenna efficiency, night effect, and the sensitivity of the receiver used aboard the aircraft.

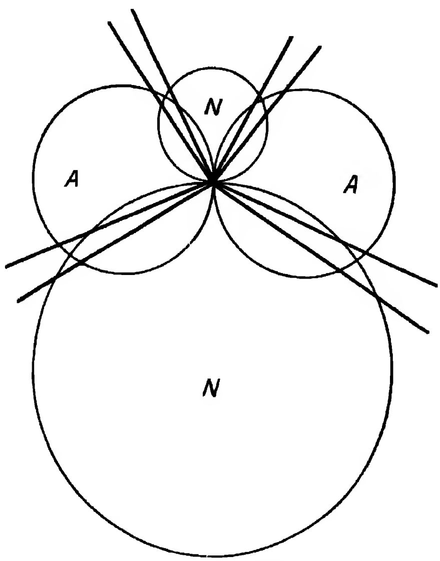



FIG. 286 Course bending

The FM (fan marker) type comprises a u-h-f transmitter having an output power of approximately 100 w and a directive-antenna system. The fan marker operates on a frequency of 75 megacycles, and its use necessitates the installation of a u-h-f receiver in the airplane in addition to the standard l-f range receiver. Fan markers are located along the airway that they serve, and, as their name implies, their radiated field pattern is in the shape of a fan, extending vertically upward from the transmitting antenna, which has its major axis at right angles to the range course and its minor axis parallel to the range course. This effect is obtained by an antenna system consisting of a horizontal dipole antenna and a semicylindrical reflector element, as shown in Fig. 287.

The signal from the FM transmitter is modulated with an audio frequency of 3,000 c. This modulation frequency is further keyed in one-, two-, three-, and four-dash groups in order that the fan marker may be identified with a particular leg of the range station. In addition to providing the aural 3,000-c tone that the pilot hears in his headset, the transmitted signal actuates a light, so that both aural and visual indication is given the pilot when his position is over the fan marker. By observing these indications, the pilot may accurately locate his position along a range course.

The Z (zone) marker comprises a u-h-f transmitter with an output power of approximately 5 w and a directive-antenna system. This type

of marker also operates on a frequency of 75 megacycles, and the same receiver may therefore be used on the aircraft for the reception of both Z marker and FM marker signals. The directive antenna system of the Z marker is so designed that a vertical, cylindrical field pattern is radiated. The antenna system is located in such a way that this radiated field pattern coincides with the cone of silence of the radio-range station. The 75-megacycle carrier of the Z marker transmitter is modulated by a continuous 3,000-c tone to provide the pilot with an aural signal and also to actuate the light indicator.

The Z marker is designed to provide a position identification of the cone of silence of the radio range and does not in itself provide the pilot

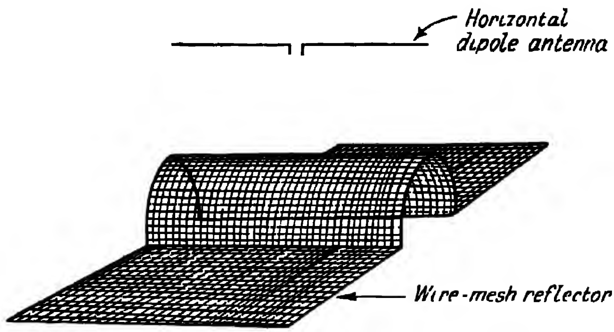



FIG. 287. Antenna system employed for fan marker stations.

with a definite indication of his position over the range station. The cone of silence must be relied on to furnish the pilot with this information. However, since the cone of silence is a *negative* indication, and since similar negative indications may occur because of a faulty receiver, a momentary failure of the transmitter power, fading, and other conditions, the primary function of the Z marker is to provide a positive means of identifying the true cone of silence of a radio range station.

Complete information regarding radio ranges, Z, FM, and M markers, weather broadcasts, and other radio facilities operated by the Civil Aeronautics Administration for the benefit and guidance of airmen is contained in "Tabulation of Air Navigation Radio Aids," published monthly by the CAA.

**The Airway Landing Beam.** An explanation of radio-range flying between radio-range stations was outlined in an earlier section of this chapter. It is possible to effect instrument approaches to airports by the use of the radio range and the sensitive altimeter alone. However, certain limitations on the use of these instruments have been recognized, and, as a general rule, instrument approaches are very seldom attempted when the ceiling is less than 300 ft and visibility less than 1 mile. This means that flight schedules are limited by weather conditions, which

must be above certain prescribed minimums. Unquestionably, the vagaries of weather conditions are the one limiting factor in absolute regularity of schedules. The air-transportation industry has long felt the need for an instrument landing system that would permit landings to be made under conditions of zero ceiling and visibility. Here again, the problem became one in which radio appears to be the only solution. The National Bureau of Standards and the old Bureau of Air Commerce of the Department of Commerce have made exhaustive tests on various systems of radio aids for instrument, or blind, landings. The first system was the landing-beam runway-localizer radio-marker system developed by the Bureau of Air Commerce. The second was an adaptation of the U.S. Army's instrument landing system which utilized a radio compass. The landing-beam system was developed at the College Park Airport, College Park, Maryland, and later tested at the Newark Airport, Newark, New Jersey. After the Civil Aeronautics Authority took over the functions previously performed by the Bureau of Air Commerce, the entire experimental and developmental program for instrument landing systems was moved to Indianapolis, Indiana, where intensive development work is constantly being conducted. The present system now in use at Indianapolis is a refinement of the original landing-beam runway localizer radio-marker system first developed by the Bureau of Air Commerce and the National Bureau of Standards.

This landing-beam system utilizes three elements the runway localizer, the radio markers, and the landing beam. The runway localizer is similar in principle to the regular radio range, except that it is on a smaller scale and gives precise and accurate directional guidance along the airport runway and the approach to the airport. By means of the runway localizer, the pilot is enabled to orient himself laterally in approaching the landing field.

The second element of the landing system comprises two low-power 75-megacycle FM-type markers that indicate to the pilot his progress toward the airport runway. These markers are a small scale version of the previously discussed 75-megacycle fan markers in common use on the airways. Reception of these markers is indicated on the aircraft by a flashing light on the instrument panel. One of the marker stations is usually located very close to the edge of the field; and the second marker is located approximately 2 miles from the edge of the field.

The third, and most interesting, element of the landing system is the landing beam, which provides the airman with vertical guidance into the airport. The landing-beam transmitter operates on approximately 90 megacycles and feeds a directive-antenna system that radiates a field pattern in the vertical plane, as shown in Fig. 288. In the horizontal plane, the beam is spread out to afford service over a sector of approx-

15° This pattern is achieved by means of an array of horizontal

half-wave dipoles arranged in a vertical sequence, as shown in Fig. 289. By proper phasing of the antenna elements (see Chap. XV), the field directly upward is canceled, and maximum radiation occurs at an angle of approximately 8° with the ground.

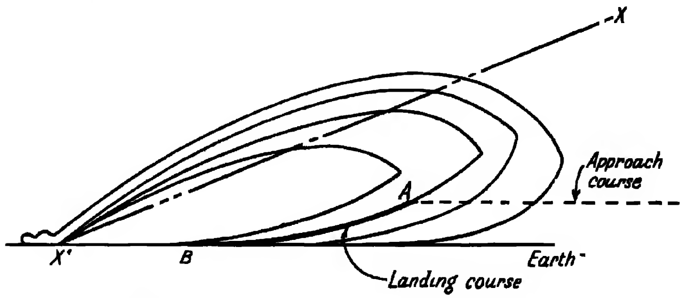



FIG. 288. Vertical-plane field pattern of glide landing beam transmitter.

In Fig. 288, if an airplane approached the airport along the line $XX'$, the signal strength would vary inversely with the distance from the transmitter, that is, the volume would increase as the airport is approached. There is no way in which this increase in signal intensity can be utilized to provide vertical guidance to the aircraft as it approaches the field, since there are any number of angles of approach that would produce a similar increase in intensity. In addition, such a straight-line approach does not coincide with the natural gliding approach of an aircraft making a landing.

The best approach to the landing field is made by causing the aircraft to follow one of the contours of equal field intensity, such as line $AB$ in Fig. 288. This path resembles the underside of an ellipsoid and very closely coincides with the natural gliding approach of a plane under normal landing conditions. The pilot is enabled to follow the course of equal field intensity by means of a visual indicator, usually a sensitive microammeter. The output of the landing-beam receiver is rectified and coupled to this meter, and the meter is so calibrated that the desired field intensity of the signal produces a deflection of the meter needle to half scale. The meter is mounted at right angles to the conventional mounting, so that the needle movement is up and down instead of from left to right. An increase in meter current causes the needle to move

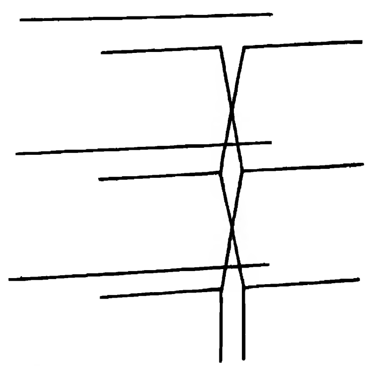


FIG. 289. Antenna array employed o produce the glide landing beam.

upward; a decrease in current causes the needle to move downward. In Fig. 288, it will be seen that any deviation of the aircraft *upward* from the chosen contour (*AB*) of field intensity will bring it into a region of higher field intensity. The meter current will therefore be greater, and the needle will move upward. Any deviation of the aircraft *downward* from the desired path will bring it into a region of lower field intensity. The meter current will therefore be smaller, and the needle will move downward. Consequently, in order to follow the gliding course, it is only necessary for the pilot to maintain the indicator needle at the half-deflection position by appropriate vertical maneuvering of his plane.

The landing beam, or vertical-glide, indicator and the runway-localizer indicator are combined in a single instrument having two movements.

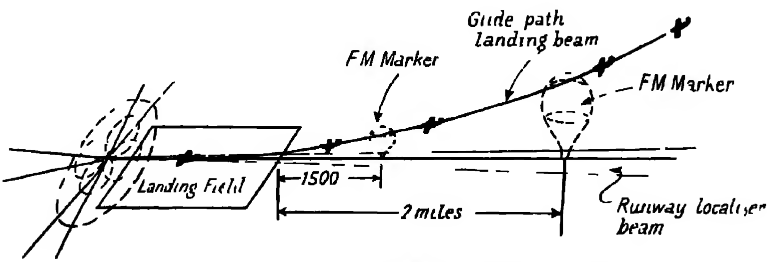



FIG. 290. Elements of an instrument landing system

The pilot is then able to check both vertical and lateral course by observation of a single instrument. The runway-localizer indicator operates on the same principle as the landing-beam indicator, except that it is mounted in the conventional horizontal position with the needle moving from left to right. Any deviation from the lateral course to left or to right is indicated on the instrument by a corresponding movement of the needle to left or right. When the proper course (both vertical and lateral) is being maintained, the needles of both movements of the instrument are at half deflection and cross in the center of the instrument. The proper point of intersection is indicated by a small circle on the instrument dial.

The following procedure is followed in effecting a landing solely on instruments (see Fig. 290). When the pilot reaches the immediate vicinity of the airport and passes directly over the radio-range transmitting station (cone of silence), his normal range receiver indicates this fact. He then tunes his localizer receiver to the frequency of the runway localizer and makes a wide circle of the field in a counterclockwise direction in order to pick up the signals of the runway localizer. He also throws a switch that places a second receiving set in operation to pick up the signals of the landing beam.

To follow the signals of the runway localizer, the pilot watches the vertical needle pointer in the previously mentioned dial on his instrument panel. Any fluctuation to right or left of the dial center shows a deviation from the true course in the indicated direction. Upon orienting himself along the runway course, generally 3 to 5 miles from the field, the pilot uses his second receiving set to pick up the signals of the landing beam for vertical guidance. From this second beam, he gets an indication on the horizontal pointer of the instrument. Without changing his lateral direction (staying on the runway-localizer beam), the airman varies altitude as necessary to bring the horizontal pointer to the center of the dial.

Continuing toward the airport, the pilot flies so that the needles cross in the circle in the center of the instrument. He is then following the center line of the course marked by the runway localizer, but with respect to the landing beam his indicator directs him along a curved line in the under part of the ellipsoidal beam.

If the aircraft dropped too far below this line, the signals received would be weak, and the needle would fall, if it climbed above this line, the needle would rise. The curved course where the signal strength remains constant brings the plane downward in a sweeping glide. The landing path, of course, is so adjusted as to clear all obstructions.

Following this unseen radio path, the airman approaches the field. About 2 miles from the edge of the field, notice is given him by a signal from a marker station on the ground below him, which is reproduced as a flash of light on his instrument panel. At this point, his altitude should be approximately 800 ft, and the pilot checks with the sensitive altimeter. Just at the edge of the field, a signal from a second marker reaches the aircraft and is again reproduced as a light flash on the instrument panel. This signal gives the pilot warning that he is near the field. If he is now at the correct altitude for this stage of the maneuver (approximately 150 ft), he continues to follow the landing beam accurately to the point where he is to make contact with the ground

In addition to the radio features of the blind-landing system, a chain of lights is utilized on the ground along the approach to the airport and down the runway itself This further facilitates the landing when the plane is close to the ground under conditions of poor visibility.

Thousands of experimental landings have been made with the various systems that have been tried during the course of the last few years. These systems include the Bureau of Standards system described in the foregoing, the German system developed by C. Lorenz, and the U.S. Army Air Corps system. The Lorenz system is essentially similar to the American (Bureau of Standards) system. In the Army system, an instrument landing is effected by use of a radio compass (visual-indicating radio-direction finder), a radio-marker receiver and a directional gyro.

After the lateral course is established by use of the radio compass on two radio stations on the line of the course, the pilot glides in on this course, checking altitude by the sensitive altimeter. After gliding down to approximately 150 ft over the second marker station, the remainder of the landing maneuver is completed with the aid of the regular flight instruments, air-speed indicator, turn-and-bank indicator, and rate-of-climb indicator.

Despite the many experimental landings that have been made with the various systems, sufficient service testing has not been conducted as yet to warrant the dispatching of aircraft to land at airports where existing weather conditions are such as to necessitate the absolute reliance on the instrument-landing system. Nevertheless, the constant research and development in this field will undoubtedly result in a system permitting safe, dependable instrument landings under zero-zero weather conditions.

**The Radio Altimeter.** Since the early days of man's first flight, pilots have recognized the desirability of an instrument that would indicate the height of the airplane above the terrain over which they were flying The present-day type of altimeter indicates altitude with relation to sea level only and is subject to the vagaries of error due to changing atmospheric pressure. Since the barometric altimeter is limited to indications of the aircraft above sea level, it does not provide indication of actual height above the earth immediately below the aircraft.

To obtain an instrument that would register actual terrain clearance, the aeronautical industry again placed its problem before the radio engineers. The radio industry has now made available a *terrain-clearance indicator*, or absolute altimeter. The present model of this instrument is a very complex assembly of radio equipment, and because of its weight, cost, and complexity, it has found little application in light aircraft. The terrain-clearance altimeter is now undergoing extensive service testing by the military and naval services and to a limited extent by the larger air lines.

This equipment is essentially a complete radio station. It includes a low-powered u-h-f transmitter and a radio receiver, a power unit, transmitting and receiving antennas, and a terrain-clearance meter that indicates the altitude in feet above the terrain.

Absolute altitude above the terrain is indicated by sending a radio wave to the ground and timing the interval required for it to reach the ground and return to the plane after it has been reflected from the ground. The output of the transmitter is radiated downward by an antenna which usually is mounted on one of the lower surfaces of the airplane. The transmitter is connected to the antenna by a coaxial transmission line. The frequency of the oscillator in the transmitter is varied, being increased and decreased automatically by a modulator.

The radio receiver is connected through another coaxial line to a second

antenna similarly attached to a lower surface of the plane. The antenna is arranged so that a minimum amount of direct wave is received from the transmitting antenna while it picks up as much as possible of the wave reflected from the ground.

The direct and reflected signals are applied to a detector circuit in the radio receiver. The frequency of the alternating current from the output of this detector is equal to the instantaneous difference in frequency existing between the direct and reflected signals (about 6 c per foot of altitude) and is directly proportional to the height of the airplane above the ground. This l-f signal from the output of the detector is amplified and applied to a frequency meter or cycle-counting circuit, including a

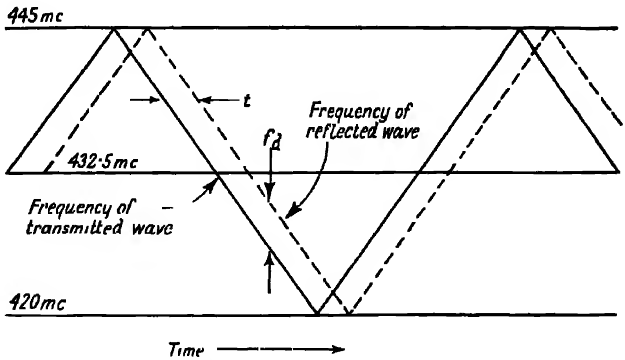



FIG. 291. Principle of operation of radio altimeter

meter with the scale calibrated in feet of altitude, which is located on the airplane instrument panel in full view of the pilot. The other units of the equipment may be located at any convenient position in the airplane.

The operation of the radio altimeter can be better understood by referring to Fig. 291. The variation of the frequency of the oscillator in the transmitter with time is indicated by the solid line. The ordinate shows the frequency of the wave transmitted to the ground for any particular instant of time.

The frequency is varied from 420 to 445 megacycles and return at the rate of 60 times per second, so that the rate of change of frequency is $3 \cdot 10^9$ c per second per second. The linear frequency variation shown in the diagram is ideal but not essential to the operation of the radio altimeter.

The broken line to the right of the solid line represents the same variation in the frequency of the wave reflected from the ground. The curve is displaced to the right by a time interval equal to the time required for the radio wave to travel to the ground and back to the

airplane or to twice the height of the airplane divided by the velocity of propagation of the radio wave Expressed mathematically,

$$\frac{2H}{C} \tag{6}$$

where $H$ = altitude of airplane above ground in miles,
$C$ = velocity of radio wave (186,000 miles per second)

The reflected wave signal is applied to the detector, together with some of the direct signal from the transmitting antenna These two signals in

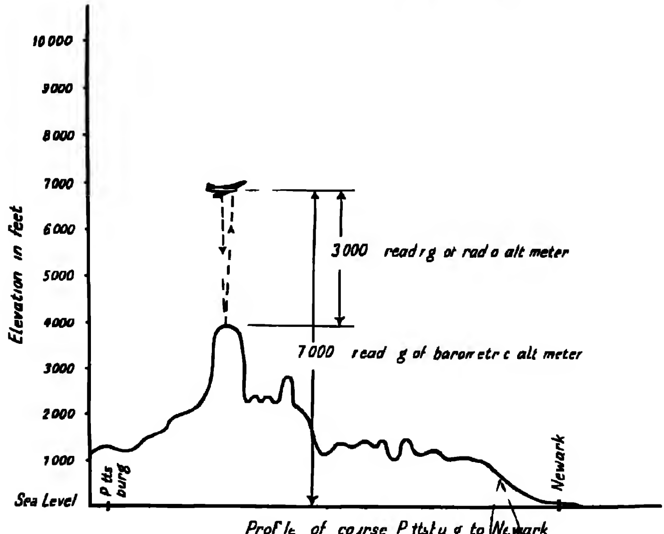



Fig 292 Application of radio altimeter

the detector produce a low beat-frequency signal $f$, which is equal to the difference in frequency between them or is equal to the product of the time delay $t$ and the rate of change of the transmitter frequency To illustrate Consider an airplane with a radio altimeter $\frac{1}{2}$ mile above the ground The time delay is

$$t = \frac{2H}{C} = \frac{2 \times 0.5}{186,000} = 0.0000054 \text{ sec} \tag{7}$$

The difference frequency from the output of the detector is

$$f_d = 3 \times 10^9 \times 0.0000054,$$
$$f_d = 16\,200 \text{ c per second,} \tag{8}$$

or, about 6 c per foot of altitude

The time delay $t$ shown in the diagram has been greatly exaggerated in comparison with the time interval of the modulating frequency in order to make the description clear

It will be noted that the difference frequency $f_d$ drops momentarily to zero twice for each complete sawtooth variation of the transmitter frequency because it is necessary to increase and then to decrease the transmitter frequency. Hence, one altitude measurement is made for each increase and decrease of the transmitter frequency, or at the rate of 120 times per second.

The introduction of the radio altimeter opens up a new field of air-navigation technique and may well be an important factor in instrument flying and landing in the future. The importance of this equipment may justifiably be compared to that of the indispensable sonic depth indicator used by the marine navigator.

At the take-off of a flight, when the airplane is only about 50 ft above the ground level of the airport, a continuous indication of altitude is shown by the terrain-clearance meter. As the altitude increases, the position of the instrument's needle changes to indicate greater distance above the ground until the altitude exceeds 5,000 ft, which is the upper limit of the instrument. Then the needle of the instrument remains off scale until the altitude exceeds 12,000 ft. In the range of 12,000 to 15,000 ft and above, the needle of the instrument may drop back momentarily on the scale because the intensity of the radio signal is low at the higher altitudes and irregular variations occur, with the result that insufficient signal is received to operate the equipment satisfactorily.

In cross-country flight, the indication of altitude above the ground will show variations according to the contour of the terrain below the airplane. This information, together with a map showing the contours of the airway sector and a reference altitude of constant level obtained from a sensitive aneroid type altimeter, may be used to locate the position of the airplane as a navigation procedure when flying above clouds or by instruments. By this method, prominent ridges and peaks of mountain ranges may be located, using the indication of the sensitive barometric altimeter as a level-flight reference.

## *QUESTIONS AND PROBLEMS**

**1.** What is indicated by the bearing obtained by the use of a bilateral radio direction finder?

**2.** Why are loop antennas, associated with radio direction finders, metallically shielded?

**3.** From how many simultaneous directions is a direction finder capable of receiving signals if adjusted to take unilateral bearings through 360°?

**4.** What is the directional reception pattern of a loop antenna?

**5.** How is the unilateral effect obtained in a direction finder?

**6.** What figure represents the reception pattern of a properly adjusted unilateral radio direction finder?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

**7.** On shipboard, what factors may affect the accuracy of a direction finder after it has been properly installed, calibrated, and compensated?

**8.** What is the principal function of a vertical antenna associated with the bilateral radio direction finder?

**9.** What is indicated by the bearing obtained by the use of a unilateral radio direction finder?

**10.** What is a compensator as used with radio direction finders, and what is its purpose?

# Chapter XVII

# MEASUREMENTS IN RADIO

Section 3.60 of the regulations of the Federal Communications Commission requires that the licensee of each standard broadcast station have in operation at the transmitter a frequency monitor independent of the frequency control of the transmitter. The frequency monitor must be of a type approved by the Commission and must have an accuracy and stability of at least five parts per million.

The licensees of ship radio stations and point-to-point radiotelegraph stations are required to provide for measurement of each operating frequency of the station and to establish a procedure for regular measurement of these frequencies. These measurements must be made by means independent of the frequency control of the transmitter and must be of an accuracy sufficient to detect deviations from the assigned frequency within one half of the authorized tolerance The tolerances imposed by law for stations in the various services are listed in Table V in the Appendix.

## *FREQUENCY MEASUREMENTS*

There are several types of frequency measuring equipment in existence, all designed to meet the various requirements of the FCC. The first of these is the *frequency monitor*, designed to give continuous visual or aural indication of transmitter-frequency deviation from the assigned frequency. The second consists of standard precision frequency measuring apparatus designed for permanent installation at a radio station. The latter type of apparatus permits making actual *measurements* of the emitted frequency at regular periods and should not be confused with the frequency monitor, which merely provides an indication of frequency deviation. The third type of frequency-measuring equipment consists of a *frequency meter*, or wave meter, designed for portable use. The latter unit is commonly used for checking ship-station frequencies where greater tolerances are permissible. Frequency meters are required to be checked often against frequency standards in order to maintain the desired precision standards set by law.

**The Frequency Monitor.** There are several ways of determining whether the frequency of an emitted carrier wave is within the required limits of the assigned frequency. One of the commonest ways is by means of a local piezo oscillator of known frequency producing a beat, or heterodyne,

with the emitted wave used in conjunction with an instrument to indicate the resultant beat frequency. The visual indicator is the only method now in common use by which it is considered that the frequency of the beat may be determined with the required degree of accuracy Approval of a frequency monitor by the FCC is based upon data taken at the Bureau of Standards. However, the Bureau of Standards does not approve or disapprove of the monitor, as this is entirely in the hands of the Commission. In approving a frequency monitor after tests at the Bureau of Standards, the Commission merely recognizes that the type of monitor has the inherent capability of functioning in compliance with FCC regulations, if properly constructed, maintained, and operated.

A block diagram of a typical frequency monitor is shown in Fig. 293. The circuit consists basically of two r-f amplifiers feeding into a common

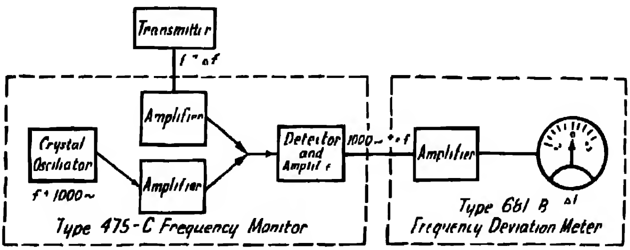



FIG 293 Block diagram of a typical frequency monitoring system (*Courtesy of General Radio Co*)

detector stage. One amplifier amplifies the transmitter carrier signal, the other, the output of a local precision crystal oscillator The precision oscillator differs in frequency from the carrier by 1,000 c and has an accuracy of plus or minus 13 parts per million

The detector output is amplified and applied to a frequency-deviation meter. This meter reads zero when the beat frequency is exactly 1,000 c (see Fig. 294) and is a conventional direct-reading a-f meter of high precision. Any deviation of carrier frequency causes a corresponding deviation of beat frequency above or below 1,000 c. The number of cycles per second deviation above or below 1,000 c is indicated directly on the meter, which can be read accurately to half a cycle.

The quartz crystal is cut to have a low temperature coefficient and is mounted in a dustproof, air-gap type of holder. The temperature coefficient is less than 2 parts per million per degree centigrade. The temperature is held to plus or minus 0.1°C by a mercury-thermostat-controlled oven, so that no appreciable variation in frequency with ambient temperature can occur. The crystal amplifier reduces the oscillator load and by isolating the oscillator prevents coupling from the transmitter from affecting the amplitude and frequency of the oscillator. Similarly, the transmitter amplifier prevents any crystal-frequency

voltage from getting into the transmitter circuits and producing 1,000-c modulation.

Loose coupling to the transmitter should be used so that the monitoring voltage can be taken directly from the transmitter-crystal control unit. This arrangement is especially important for transmitters using low-level modulation.

Most frequency-monitor units are equipped with auxiliary meters to

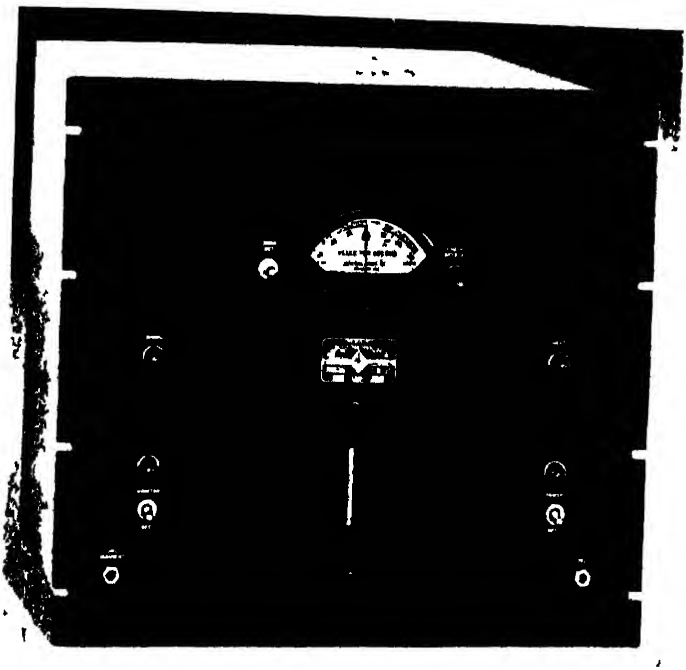


FIG. 294 Frequency deviation monitor (*Courtesy of General Radio Co.*)

furnish an indication of correct signal level from the transmitter and from the crystal oscillator The monitor amplifier gain can be varied to adjust the input according to the desired levels as indicated on the level indicating meters and thus ensures correct coupling when the monitor is installed and prevents false readings due to overloading.

The a-f amplifier incorporates a-v-c control. By using a large delay voltage, constant input voltage to the frequency deviation meter is held over a beat-frequency amplitude range of 0.5 to 8 v. The monitor accuracy is therefore not subject to variation caused by line-voltage fluctuation, changes in transmitter adjustment, and aging of tubes.

**Frequency Measuring.** Rough frequency measurements using portable equipment, such as for ship-station measurements, are made with heterodyne-frequency meters. A heterodyne-frequency meter is essentially a calibrated variable oscillator. Commercial heterodyne-frequency meters are accompanied by numerous graphs giving extensive calibrations for various dial settings within the range of the instrument. Most meters of this type utilize electron-coupled oscillators because of the good frequency stability and rich harmonic output. The oscillator circuits are completely shielded so that external circuit coupling, which might conceivably affect oscillator frequency, is eliminated.

The average well-designed commercial heterodyne-frequency meter

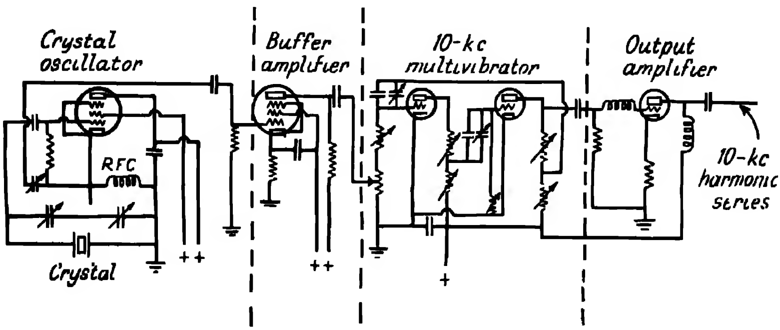



FIG. 295 Schematic diagram of secondary frequency standard

will maintain an accuracy of 0.1 per cent over comparatively long periods of time, provided that corrections are made for variations of temperature. Manufacturers usually provide temperature correction factors for normal temperature ranges. If the frequency meter is often checked against a standard, accuracies of 10 to 15 parts per million over short periods of time may be attained.

Commercial heterodyne frequency meters incorporate a detector stage utilized to produce beat frequencies or heterodynes between the meter oscillator and the signal to be measured. The procedure involved in making a measurement is comparatively simple. The frequency meter is coupled loosely to the transmitter whose frequency it is desired to measure. A short length of wire connected to the instrument input terminal is usually sufficient to pick up the transmitter signal. The heterodyne oscillator is then tuned to zero beat with the signal as indicated audibly in a pair of headphones connected to the meter detector output circuit. With the modern instruments now available, it is possible to accomplish the zero-beat adjustment within half a cycle. Once this adjustment has been attained, the frequency as obtained from the calibration for this dial setting will be the frequency of the signal being measured.

More accurate measurements of frequency are obtained by the use of a *standard frequency assembly*. Such an assembly consists of a receiver, a heterodyne frequency meter, a high-precision frequency standard (signal generator), a heterodyne detector, an interpolation oscillator, and some type of indicating device, either earphones or a visual frequency meter for very low frequencies.

The frequency standard, often called the **comparison oscillator,** since the signal being measured is "compared" with, or checked against, this standard, consists of a precision crystal oscillator, buffer amplifier, 10-kc multivibrator, and output amplifier. A circuit of a typical frequency standard as used in standard frequency assemblies is shown in Fig. 295.

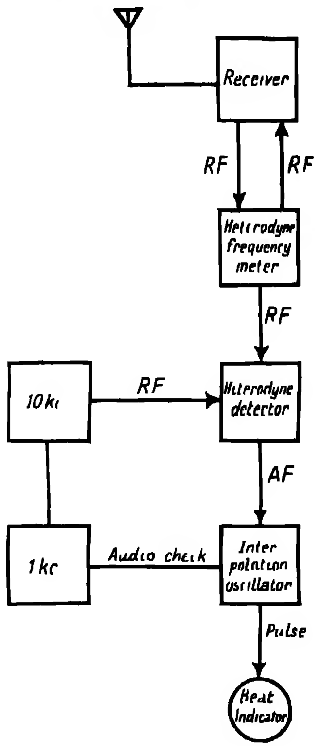



FIG. 296. Basic steps in conducting a frequency measurement.

Several features of the crystal oscillator which contribute toward frequency stability are worthy of note. The crystal is kept free from temperature-variation effects by a thermostatically controlled oven. The dustproof air gap type of crystal holder introduces very little restraint on the vibration of the crystal, permitting it to vibrate freely. Some commercial models eliminate detrimental effects from resonances in the air by mounting baffle plates a quarter wave length from the ends of the crystal.

The buffer amplifier effectively isolates the crystal stage from succeeding stages and eliminates any coupling effects that might impair the operation of the oscillator.

The multivibrator develops a great number of harmonics of the crystal frequency Commercial instruments usually utilize a 10-kc sequence, so that any two adjacent harmonics are 10 kc apart. Since a single well-designed multivibrator will develop harmonics up to the five-hundredth, a single comparison oscillator is usable for measuring frequencies over a wide range. Each of the harmonics will have the same degree of accuracy as the crystal. The output amplifier provides a convenient means of coupling to the heterodyne detector load.

The procedure involved in conducting a measurement is as follows. The signal to be measured is tuned in on the receiver and transferred to the heterodyne frequency meter as discussed in the previous section.

The heterodyne frequency meter is then coupled to the heterodyne detector as is also the harmonic output of the comparison oscillator (frequency standard). Refer to the block diagram of Fig. 296. Since the harmonics of the comparison oscillator are separated by 10 kc, the signal must be within 5 kc of one of these harmonics. An a-f beat note of 5 kc or less will therefore be developed in the heterodyne detector as a result of the interaction of the signal (from the heterodyne frequency meter) and one of the comparison oscillator harmonics. When the harmonic frequency is identified, the remaining procedure consists of identifying the a-f beat note. The frequency of the signal is then equal to the harmonic frequency plus or minus the beat frequency, depending upon whether it is a higher or lower harmonic. The beat frequency is identified by feeding it into an *interpolation oscillator*. An interpolation oscillator is simply a calibrated variable a-f oscillator having a range from 0 to 5 kc, since 5 kc is the maximum amount by which the signal can deviate from a comparison oscillator harmonic. The interpolation oscillator is tuned to zero beat with the beat note fed into it from the heterodyne detector. The frequency of the beat note can then be taken directly from the interpolation oscillator dial reading. This frequency, as mentioned before, is added to or subtracted from the proper comparison-oscillator harmonic to obtain the signal frequency.

It is apparent that the accuracy of the measurement is dependent not only upon the accuracy of the comparison oscillator, but also upon that of the interpolation oscillator. Interpolation oscillators are now available, permitting dial readings of approximately 1 c per dial degree. Many installations utilize a 1-kc sequence output from the frequency standard in addition to the 10-kc sequence. The 1-kc sequence output is fed into the interpolation oscillator and is checked against dial readings for calibration purposes.

Usually, very little difficulty is encountered in distinguishing which harmonics of the comparison oscillator are being used, since the frequency of the signal being measured is always approximately known. Thus, if the signal being measured has a frequency in the vicinity of 1,896 kc, it is apparent that it lies between the 189th and 190th harmonic of the comparison oscillator. If difficulty is experienced in identifying the harmonic that is being used, the heterodyne frequency meter should be increased in frequency slightly after the interpolation-oscillator reading has been obtained. If increasing the frequency of the heterodyne frequency meter results in an *increase* in the a-f beat note, it is evident that the *lower* harmonic is being used. The beat-note frequency obtained from the interpolation oscillator must then be *added* to the lower harmonic to obtain the signal frequency. If increasing the frequency of the heterodyne frequency meter results in a *decrease* in the a-f beat note, the heterodyne-frequency meter is beating with the *higher* harmonic. The beat-note

frequency should then be *subtracted* from the higher harmonic to obtain the signal frequency.

A consideration of the detailed procedure involved in measuring a specific frequency will clarify the operations to be performed. Assume that it is desired to measure the frequency of a signal that is supposed to be on 1,896 kc. The signal is tuned in on a receiver and transferred to the heterodyne-frequency meter as outlined above. Together with the output of the comparison oscillator, the frequency meter is coupled to the heterodyne detector. From the roughly known location of the frequency, it is known that the signal lies between the 189th and 190th harmonics of the oscillator. The interpolation oscillator is tuned to zero beat with the beat note produced in the detector and is found to read 4,000 c. *After* this latter measurement has been made, the frequency of the heterodyne frequency meter is increased slightly, with the result that the beat note *decreases* slightly. The 4,000 c is therefore subtracted from the 190th harmonic (1,900 kc) giving the frequency as 1,896 kc. This is therefore the frequency of the signal being measured.

The frequency of an unknown signal can always be roughly identified by the approximate calibration of the receiver used or by the calibration of the heterodyne frequency meter. In cases where even this identification is uncertain, such as on the higher frequencies where adjacent harmonic differences may be small, the customary procedure is to modulate certain harmonics of the comparison oscillator. The modulation is accomplished by operating a 100-kc multivibrator from the same basic comparison oscillator source. The output of this multivibrator is modulated with a distinctive tone and coupled with the 10-kc multivibrator to the heterodyne detector. Every tenth harmonic of the 10-kc multivibrator will then be marked, since it beats with the modulated 100-kc harmonics. Intermediate 10 kc harmonics can be identified by counting backward or forward to the nearest modulated harmonic.

This procedure limits the identification of a frequency approximation to within 100 kc, the difference between two modulated harmonics. It is very rarely that a frequency cannot be approximately identified within this wide range without the use of instruments. Once the rough frequency identification has been made, the exact frequency can be accurately measured as outlined above.

## *ANTENNA RESISTANCE MEASUREMENTS*

In Chap. XV it was shown that the value of effective resistance of an antenna is a very useful factor to have at hand. The effective resistance must be known in order to design a matching system properly for a nonresonant transmission line. When computing the power put into an antenna, the effective resistance of the antenna is an important constant

which must be included in the calculations. Furthermore, when the effective resistance is known, a *dummy antenna* can be constructed having the same resistance. Such a dummy, as a nonradiating load, can be coupled to the transmitter when conducting experiments or running tests

There are two standard methods of measuring antenna resistance, namely, the *resistance-variation method* and the *half-deflection method*. The former is usually considered more accurate, but both methods are used with success. As in any other type of measurement, the results depend to a large extent upon the care and efficiency with which the experiment is conducted.

**The Resistance-variation Method.** Owing to the behavior of electricity at high frequencies, measurements of circuit constants for high frequencies are not so easily obtainable as are those at the lower frequencies. Ordinarily, to obtain the resistance of an a-c circuit, the current and voltage would be measured at resonance and Ohm's law applied to obtain $R$. In an antenna system, it is found that the current is the only factor that can easily be measured However, by measuring the antenna current at various artificially introduced resistances of the circuit, several pairs of simple simultaneous equations can be obtained from which the antenna resistance can readily be computed

In any a-c circuit, according to Ohm's law,

$$I = \frac{E}{Z}. \tag{1}$$

In any a-c circuit at resonance,

$$Z = R, \tag{2}$$

and Eq. (1) becomes

$$I = \frac{E}{R} \tag{3}$$

Rearranging,

$$E = IR \tag{4}$$

If a known value of resistance $R_1$ is added in series with the antenna system, a new value of current $I_1$ will be obtained, and according to Ohm's law

$$I_1 = \frac{E}{R + R_1} \tag{5}$$

Rearranging,

$$E = I_1R + I_1R_1. \tag{6}$$

Since the value of $E$ is the same in both cases, Eqs (4) and (6) may be combined (solution by comparison), with the result that

$$IR = I_1R + I_1R_1. \tag{7}$$

Then,

$$IR - I_1R = I_1R_1, \tag{8}$$

and

$$R(I - I_1) = I_1R_1. \tag{9}$$

Dividing by $(I - I_1)$,

$$R = \frac{I_1R_1}{I - I_1}, \tag{10}$$

$$R = \frac{R_1}{\frac{I}{I_1} - 1}. \tag{11}$$

This method, known as the **resistance-variation method,** is based on Eq. (11), where

$R$ = unknown antenna resistance,
$R_1$ = value of added resistance;
$I$ = current with no resistance added;
$I_1$ = current with added resistance.

Several values of added resistance $R_1$ should be used and the results averaged for greater accuracy. Every effort should be made to use

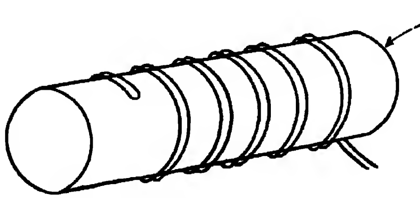



FIG. 297 Noninductive wire wound resistor.

resistors that are, so far as possible, noninductive. Other than the resistors, no additional apparatus is necessary to perform the measurement. A drawing of a noninductive resistor is shown in Fig. 297. Although actual zero inductance cannot be obtained, a resistor wound in the fashion illustrated will closely approach the ideal. Some provision should be made to facilitate opening of the antenna circuit to provide for the insertion of the resistors.

The experiment is conducted as follows. The antenna circuit is tuned to resonance with the transmitter and the current $I$ carefully noted. A known value of resistance $R_1$ is then inserted in series and the current $I_1$ again carefully measured. The values of $I$, $I_1$, and $R_1$ obtained are then substituted in Eq. (11). The experiment is then repeated several times, each time using different values of $R_1$ and substituting in Eq. (11). The various values of $R$ obtained should be averaged for the final value. If

the measurements are carefully conducted, very little discrepancy will be found to exist between the different $R$ values obtained.

Care should be taken during the measurements to maintain a constant transmitter voltage output, since any fluctuation in the voltage transferred to the antenna will introduce serious error in the results. It is best to check plate current and voltage of the last transmitter stage before each antenna-current reading and to correct any existing discrepancy. It is advisable to use an r-f milliammeter to measure the antenna current and to adjust the transmitter output so that the current falls well within the meter scale before beginning measurements. If necessary, a shunt can be used across the meter, but it will necessitate correcting the current readings before substitution in the resistance equation.

The antenna circuit *must* be tuned to absolute resonance with the transmitter before each current reading is taken. Any deviation from the resonance condition will cause the current to be a function of the antenna impedance rather than of the antenna resistance, and the reading will be erroneous so far as the purpose of the experiment is concerned. Such deviation from resonance may be caused by the introduction of inductance or capacitance into the circuit when the resistors are inserted. The resonance adjustment must therefore be checked whenever a resistance is added to the circuit and just before each meter reading.

**The Half-deflection Method.** Another somewhat simpler method of measuring the antenna resistance is by means of the so-called **half-deflection method.** This method is also based on Ohm's law for the resonant condition $E = IR$.

It is apparent from this simple equation that, if the voltage $E$ is kept constant, current $I$ will vary inversely as the resistance $R$. In other words, if $E$ is kept constant and $R$ is doubled, then $I$ will be halved.

The apparatus is set up for this measurement exactly as in the case of the resistance variation method. The driver, or transmitter, is tuned to resonance with the antenna, and an accurate measurement of the antenna current is made. The antenna circuit is then opened, and enough resistance is added to cause the antenna current to fall to half its former value. When this condition is obtained, the value of the added resistance will be exactly equal to the antenna, or unknown, resistance. Great care should be taken to maintain a constant voltage input to the antenna circuit, since any variation in $E$ will destroy the accuracy.

In performing this measurement, as in the resistance-variation measurement, the resistor used should be of the noninductive type. Whatever small change in tuning is necessary to regain the condition of resonance at various stages of the experiment should be made.

## *ANTENNA INDUCTANCE AND CAPACITANCE MEASUREMENTS*

The mathematical treatment of currents in circuits having distributed inductance and capacitance (such as antenna circuits) is generally concerned with the theoretical case wherein these quantities are uniformly distributed. Because of end effects, this condition cannot be strictly obtained in antenna circuits, although it is closely approached in transmission lines terminated in their characteristic impedances. It is evident, therefore, that the distributed inductance and capacitance of an antenna system are values that are not easily computed. These values, however, can be measured for a particular antenna system. When one value has been found, it is possible to compute the other very simply. A method of measuring the inductance of an antenna system is outlined below.

The inductance measurement is based on the familiar equation

$$\text{wave length} = 1.884\sqrt{LC}, \tag{12}$$

where wave length is in meters, $L$ is inductance in microhenrys, and $C$ is capacitance in micromicrofarads.

The apparatus necessary to perform the experiment includes one or two precision inductances of accurately known value, a signal generator or variable oscillator of some sort, and the usual frequency-measuring equipment.

In proceeding with the measurement, it is assumed that the normal operating frequency at which the antenna is resonant is known and has been measured accurately.

One of the precision inductances is connected in series with the antenna system. The signal generator is then tuned to resonance with the antenna system (with inductance inserted), and the frequency of the signal generator is measured by the means previously described in this chapter. When this frequency is known, the wave length is easily computed and substituted in Eq. (12), which is modified as follows:

$$\lambda_1 = 1.884\sqrt{(L_1 + L_a)C}, \tag{13}$$

where $L_1$ = known inductance inserted in microhenrys;
$L_a$ = unknown antenna inductance in microhenrys;
$C$ = antenna capacitance in micromicrofarads.

The experiment is then repeated, using a different value of known inductance $L_2$. The wave length $\lambda_2$ obtained from the frequency measurement this time is again substituted in Eq. (12), with the result that

$$\lambda_2 = 1.884\sqrt{(L_2 + L_a)C}. \tag{14}$$

A pair of simultaneous quadratic equations are now available, which can be solved for $L_a$ as follows:

$$\lambda_1 = 1.884\sqrt{(L_1 + L_a)C}, \tag{15}$$

$$\lambda_2 = 1.884\sqrt{(L_2 + L_a)C}. \tag{16}$$

Squaring both sides of both equations,

$$\lambda_1^2 = (1.884)^2(L_1 + L_a)C, \tag{17}$$

$$\lambda_2^2 = (1.884)^2(L_2 + L_a)C. \tag{18}$$

Dividing Eq. (17) by $(L_1 + L_a)$,

$$\frac{\lambda_1^2}{L_1 + L_a} = (1.884)^2C. \tag{19}$$

Dividing Eq. (18) by $(L_2 + L_a)$,

$$\frac{\lambda_2^2}{L_2 + L_a} = (1.884)^2C. \tag{20}$$

Equating Eqs. (19) and (20),

$$\frac{\lambda_1^2}{L_1 + L_a} = \frac{\lambda_2^2}{L_2 + L_a}. \tag{21}$$

By cross products,

$$\lambda_2^2L_1 + \lambda_2^2L_a = \lambda_1^2L_2 + \lambda_1^2L_a. \tag{22}$$

Rearranging,

$$\lambda_2^2L_a - \lambda_1^2L_a = \lambda_1^2L_2 - \lambda_2^2L_1, \tag{23}$$

or

$$L_a(\lambda_2^2 - \lambda_1^2) = \lambda_1^2L_2 - \lambda_2^2L_1, \tag{24}$$

and

$$L_a = \frac{\lambda_1^2L_2 - \lambda_2^2L_1}{\lambda_2^2 - \lambda_1^2}. \tag{25}$$

Equation (25) is the ultimate formula upon which the measurement is based, and the values of wave length and inductance used in the experiment should be substituted directly in this equation. The derivation of the formula is given for those interested and need not enter into the calculations for determining antenna inductance. All values of $L_a$, $L_1$, and $L_2$ are in microhenrys.

It should be understood that in the above measurement the inductance should be inserted directly in the radiating portion of the antenna itself and not in the feeder lines if the antenna is fed by some sort of transmission line. Adding inductance to the coupling coil at the station end of a transmission line will result only in lowering the efficiency of coupling to the transmitter output and will not affect the antenna constants in any way whatsoever.

Once the inductance of an antenna has been ascertained by this measurement, it is a simple matter to compute the capacitance. The value of inductance found from Eq. (25) is substituted in Eq. (12). In this case, the wave length used should be the wave length of the antenna system without any added inductance. In other words, the wave length of the antenna system as it is normally used should be the value substituted in Eq. (12). Equation (12) is then rearranged as follows.

$$C = \frac{\lambda^2}{(1.884)^2 L}, \tag{26}$$

where $C$ = antenna capacitance in micromicrofarads;
$L$ = antenna inductance in microhenrys.

## *FIELD STRENGTH MEASUREMENTS*

The field strength of a radio wave at a distance from the transmitting antenna is determined by measuring the voltage that the wave induces in a receiving antenna. Such measurements are of great value in determining the primary and secondary service areas of broadcast stations, in choosing sites for receiving stations, and in determining interference areas existing between two stations. Field-intensity measurements are also often required by the FCC for presentation in support of applications or evidence at hearings before the Commission. Theoretical calculations to establish the performance of directional antennas are also required by law to be checked by actual field-intensity measurements of the signal radiated by such antennas.

**General Principle of Field-intensity Measurements.** There are several approved methods of making field intensity measurements. In general, all consist of picking up the signal to be measured on a sensitive receiver and comparing the receiver output with this signal to the receiver output when a signal of known voltage from an accurately calibrated oscillator is introduced into the receiver.

It is apparent that a number of factors enter into the calculations. The over-all gain of the receiver, the type and gain of the antenna used, and the gain of the receiver input circuits must be considered. The standard measuring procedure is as follows: The signal is tuned in on a standard field-intensity superheterodyne receiver. Such receivers are equipped with a microammeter in the second-detector output, a calibrated attenuator in the i-f amplifier, and a number of switching arrangements to facilitate performing the various operations of the measurement outlined below.

When the signal is tuned in, the calibrated i-f attenuator is adjusted to a value of attenuation $a$, which produces a convenient deflection of the output circuit microammeter. By means of a switch in the loop-antenna

circuit, the local comparison oscillator is coupled directly to the loop antenna. The output of the oscillator is adjusted until an arbitrarily chosen voltage $E$ appears on the grid of the first detector. Standard field-intensity meters are equipped with a switching arrangement that permits converting the first-detector tube into an electronic voltmeter with a microammeter in the plate circuit. The microammeter is calibrated in volts of the grid input circuit.

The i-f attenuator is now adjusted to whatever value $b$ is required to produce the second-detector output the same as that of the original signal. The actual voltage produced by the signal at the first-detector grid can now be computed from the relation of the attenuation values $a$ and $b$. Expressed mathematically,

$$E_s = \frac{E_o b}{a}, \tag{27}$$

where $E_s$ = signal voltage at first-detector grid,
$E_o$ = oscillator voltage at first-detector grid,
$a$ = i-f signal attenuation required for given output;
$b$ = i-f oscillator attenuation required for same output.

This process of checking the signal against the oscillator is called **calibrating the receiver,** since it amounts to a calibration of the receiver gain. In the process of making intensity measurements over comparatively long periods of time, the readings of the second-detector output microammeter are recorded, and periodic calibrations are made with the comparison oscillator to ensure uniform gain of the receiver over the time interval involved. Many commercial field-intensity meter installations are equipped with recording microammeters or milliammeters. Such instruments provide a continuous recording or track of the output level.

The procedure outlined above makes it possible to calculate the signal voltage *at the first-detector grid*. It does not, however, allow for the gain in the preceding r f circuits and in the antenna itself. Since the field strength that it is desired to ascertain is the actual voltage induced in the antenna itself, further corrections must be applied to derive the antenna signal voltage from the first detector grid signal voltage. The relation between these two voltages is obtained as follows: Without disturbing the comparison oscillator output, the oscillator is coupled, by means of a switching arrangement inherent in the apparatus, directly to the first-detector grid instead of to the antenna. The i-f attenuator is again adjusted to whatever new value of attenuation $c$ is required to bring the output to the previous level. The difference between antenna signal voltage and first-detector grid voltage is then the same as the attenuation ratio of $c$ to $b$. The voltage of the oscillator when coupled to the antenna is then $b/c$ times the voltage of the oscillator at the grid. Since the gain, or possibly loss, of the antenna and r-f input circuits is

a fixed value, this same relation holds true for the signal voltage. In other words, the voltage developed by the signal on the antenna is $b/c$ times the voltage developed on the first detector grid by the signal. Since the signal voltage developed at the first-detector grid is given by Eq. (27), it follows that

$$E = \frac{E_o b}{a} \cdot \frac{b}{c}, \tag{28}$$

$$E = \frac{E_o b^2}{ac}, \tag{29}$$

where $E$ = signal voltage developed in the antenna;
$E_o$ = oscillator voltage at first-detector grid;
$a$ = i-f attenuation required for given signal output;
$b$ = i-f attenuation required for given oscillator output (oscillator coupled to antenna);
$c$ = i-f attenuation required for given oscillator output (oscillator coupled to first-detector grid).

When a continuous graphic recording instrument is not used, it is necessary to make notations of the output readings at predetermined intervals. Often such readings are utilized to make a graph of the measurements. The calibration should be checked frequently, the length of the intervals between calibration depending upon the stability of the instrument. If, during the course of calibration, it is necessary to change the receiver gain, notations of the new attenuation settings should be made at the proper place in the output reading notations, to ensure that the proper attenuation values ($a$, $b$, and $c$) are used in conjunction with the proper output reading values when making the final calculation with Eq (29).

One of the difficulties encountered when making field intensity measurements is that caused by fading of the signal. The customary procedure is to take output readings at the peaks of successive fast fades and, by averaging these values over an interval of time to get the longer fade characteristic. Of course, if a graphic recording instrument is used, the necessity for this operation is obviated.

## *ANTENNA POWER MEASUREMENTS*

The *maximum rated carrier power* of a standard broadcast station has been defined by the FCC as the sum of the applicable power ratings of the vacuum tubes employed in the last stage of the transmitter. Only vacuum tubes of approved ratings may be used in the last radio stages of such transmitters, and such approved ratings are given only upon submission of the ratings by the tube manufacturers. The approved ratings for various vacuum tubes for this usage are fixed by the Commission and a tabulation of these ratings is set forth in the Commission's

"Standards of Good Engineering Practice concerning Standard Broadcast Stations."

The actual **operating power** of a standard broadcast station may be defined as the r-f power put into the antenna. According to the FCC "Rules and Regulations," operating power may be determined by two methods—*indirect measurement* and *direct measurement.*

**Indirect Measurement of Operating Power.** The indirect measurement of operating power of a standard broadcast station consists of applying an efficiency factor to the input power to the final radio stage. The input power to the last stage is the product of the plate voltage $E_p$ and the total plate current $I_p$, or

$$\text{input power} = E_p I_p. \tag{30}$$

The operating power, representing the actual power put into the antenna, is derived from the formula

$$\text{operating power} = E_p I_p F, \tag{31}$$

where $F$ is an efficiency factor determined by the FCC and depends upon the maximum rated carrier power of the transmitter, the class of amplification used, the type of modulation used, or the approved power rating of the tubes used in the final stage. The efficiency factors for various classes of broadcast transmitters are outlined below.

Stations of *all* powers utilizing low-level modulation with the final power-amplifier stage operating class B have an efficiency factor of 0.35 as determined by the FCC.

An efficiency factor of 0.65 has been fixed by the FCC for stations of *all* powers utilizing low-level modulation with linear operation of the final power amplifier where the efficiency approaches that of class C.

The efficiency factors for stations of all powers utilizing grid modulation in the final radio stage has been fixed by the FCC according to the approved power rating of the tubes used in this stage. An efficiency factor of 0.25 has been fixed by the Commission for all transmitters in this category which utilize vacuum tubes as listed in Table C of the "Approved Power Rating of Vacuum Tubes," a tabulation listed in the FCC "Standards of Good Engineering Practice concerning Standard Broadcast Stations." The efficiency factor has been fixed at 0.35 if the types of tube used are listed in Table D of the same publication.

The efficiency factor for stations utilizing plate modulation in the final power-amplifier stage has been determined by the FCC in accordance with the maximum rated carrier power of the transmitter. If the maximum rated carrier power is between 100 and 1,000 w, the efficiency factor is fixed at 0.70. If the maximum rated carrier power is 5,000 w or greater, the efficiency factor is fixed at 0.80.

In computing operating power by indirect measurement, the above

factors apply in all cases. No distinction is made when the operating power is less than the maximum rated carrier power.

**Direct Measurement of Operating Power.** The operating power, or antenna input power, is determined directly by measuring the antenna current. The power is derived from the familiar equation

$$P = I^2R, \tag{32}$$

where $P$ = operating power;
$I$ = antenna current;
$R$ = antenna resistance at the point where the current is measured.

The measurement is, of course, conducted at the operating frequency. Direct measurement of the antenna input power will be accepted as the operating power of the station provided that the data on the antenna-resistance measurements are submitted under oath and detailed descriptions given of the method used and the data taken. The antenna current must be measured by an ammeter of accepted accuracy pursuant to the section of the FCC regulations concerning indicating instruments

Measurements of antenna resistance made by any standard method will be accepted by the FCC provided that satisfactory evidence is submitted in accordance with the FCC "Standards" as to the procedure used, accuracy of instruments, and qualifications of the engineer conducting the measurements. The resistance-variation method and substitution method discussed in an earlier part of this chapter are both acceptable.

Standard broadcast stations permitted to determine the operating power by the indirect method and to employ greater daytime power than nighttime power must maintain the same operating efficiency for both daytime and nighttime operation. In order to determine whether the same operating efficiency obtains, the following procedure is used.

The apparent antenna resistance is computed from the daytime (highest power) operating constants. The nighttime power in the antenna is then computed by the $I^2R$ method, using the apparent resistance previously determined. If this computed antenna power agrees with the nighttime operating power determined by the indirect method within plus or minus 5 per cent, the station is considered to be complying with the requirement of maintaining the same operating efficiency. In case the antenna current is subject to variations due to weather or other conditions, an attempt should be made to arrive at an average value for the purpose of these computations.

## *MODULATION MEASUREMENTS*

The licensee of a broadcast station is not authorized to operate the transmitter unless it is capable of delivering satisfactorily the authorized

power with a modulation of at least 85 per cent. When the transmitter is operated with 85 per cent modulation, not over 10 per cent combined a-f harmonics is permitted to be generated by the transmitter. The term **combined audio harmonics** is defined by the Commission as the arithmetical sum of the amplitudes of all the separate harmonic components.

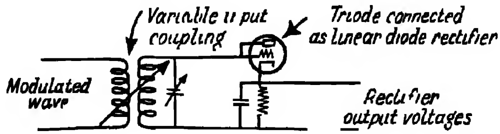



FIG. 298. Linear rectifier circuit employed to obtain modulated carrier-voltage values

The percentage of modulation is a measure of the amount that the carrier wave is changed in amplitude when an a-f wave is superimposed upon it. Expressed mathematically,

$$M = \frac{E_{max} - E_{min}}{2E_{avg}} \times 100, \tag{33}$$

where $M$ = percentage of modulation,
$E_{max}$ = maximum (peak) value of carrier voltage,
$E_{min}$ = minimum (trough) value of carrier voltage;
$E_{avg}$ = average value of carrier voltage (value of carrier voltage with no modulation)

The percentage of modulation is defined by the FCC as the average of the positive and negative modulation percentages.

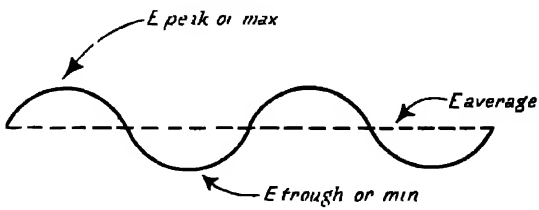



FIG. 299 Wave shape of linear rectifier output voltage

The maximum, minimum, and average carrier voltages are obtained by feeding a small portion of the transmitter output to a linear rectifier, such as the diode circuit of Fig. 298. A small coil of a few turns placed close to the output tank circuit is usually sufficient, since an r-f voltage of 20 to 30 v is necessary. The diode rectified output is passed through an r-f filter and the remaining d c and a-f components are applied to an electronic voltmeter of the peak-indicating type A typical curve of the diode output when the transmitter is modulated by a pure sine-wave a-f current is shown in Fig. 299.

The circuit of a *peak-indicating type of voltmeter* is shown in Fig. 300(a). The bias voltage as read on voltmeter $V$ is adjusted by means of the potentiometer until current just begins to flow in the circuit, as indicated by microammeter $M$. This bias voltage is then approximately equal to the peak value of the diode a-c component, which is the voltage being measured. If the bias is slightly less than the peak input voltage, the plate will always be negative, and no current will flow at any time.

The minimum carrier voltage is measured by means of a *trough*, or *negative-crest, indicating type of voltmeter*. The circuit is shown in

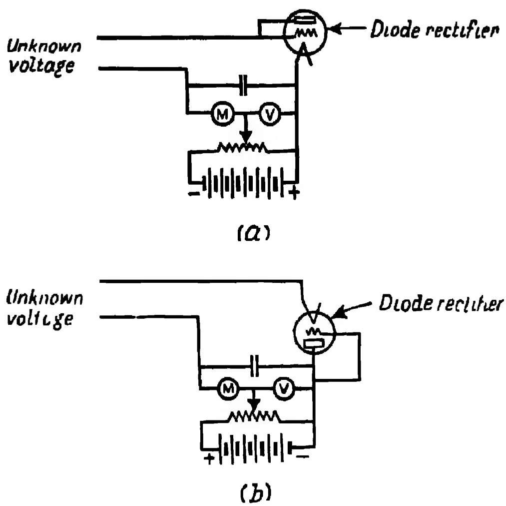



Fig. 300. Rectifier types of voltmeters (a) Peak (b) Negative crest

Fig. 300(b). The principle of operation here is the reverse of the peak-indicating meter. The potentiometer is adjusted so that plate current flow is barely perceptible as indicated on the microammeter. The value of bias voltage as read on voltmeter $V$ is then approximately equal to the minimum, or trough, voltage of the a-c component under measurement.

The average value of carrier voltage can be ascertained by the d-c voltage across the load resistor of Fig. 298.

All the foregoing voltages, when ascertained by this method, are relative values and do not represent the actual values of carrier voltage. Nevertheless, the relations between them are the same as the relations between the actual maximum, minimum, and average values of carrier voltage. They may therefore be substituted directly in Formula (33) to calculate percentage modulation. Where the actual carrier values are desired, it is necessary to calibrate the equipment for the particular

installation and to take into account transmitter output, coupling efficiency, and so on.

Percentage modulation may also be determined by direct observation on a cathode-ray oscilloscope. This instrument can be operated directly from the modulated wave, thus reducing to a minimum the possibility of error, although the width of the luminous line limits the accuracy of observation. The oscilloscope has the further disadvantage that access to the modulating voltage must be had in order to obtain a suitable sweep voltage. This instrument has the advantage, however, that rough indications of nonlinearity and audio distortion may be obtained very quickly.

### *QUESTIONS AND PROBLEMS**

**1.** If a wave meter having an error proportional to the frequency is accurate to 20 c when set at 1,000 kc, what is its error when set at 1,250 kc?

**2.** What is the meaning of zero beat as used in connection with frequency-measuring equipment?

**3.** Describe the technique used in frequency measurements employing a 100-kc oscillator, a 10-kc multivibrator, a heterodyne frequency meter of known accuracy, a suitable receiver, and standard frequency transmission.

**4.** What factors enter into the determination of power of a broadcast station employing the indirect method of measurement?

**5.** What is the device called that is used to derive a standard frequency of 10 kc from a standard-frequency oscillator operating on 100 kc?

**6.** If a heterodyne frequency meter having a straight line relation between frequency and dial reading has a dial reading of 31.7 for a frequency of 1,390 kc and a dial reading of 44.5 for a frequency of 1,400 kc, what is the frequency of the ninth harmonic of the frequency corresponding to a scale reading of 41.2?

**7.** Describe a method of determining antenna resistance.

**8.** If a heterodyne frequency meter having a calibrated range of 1,000 to 5,000 kc is used to measure the frequency of a transmitter operating on approximately 500 kc by measurement of the second harmonic of this transmitter and the indicated measurement was 1,008 kc, what is the actual frequency of the transmitter output?

**9.** If a frequency meter having an over-all error proportional to the frequency is accurate to 10 c when set at 600 kc, what is its error in cycles when set at 1,110 kc?

**10.** What is a multivibrator, and what are its uses?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

# Chapter XVIII

# STUDIO AND CONTROL EQUIPMENT

In addition to the transmitter proper and the associated modulator and speech-amplifier equipment a number of other types of specialized apparatus are required to complete the complement of broadcast station equipment. Devices are required to supply the necessary amplification to the a-f signal before it reaches the transmitter speech amplifier. Additional units are required to adjust the frequency response of transmission lines, to couple numerous circuits to a common input circuit efficiently, and to provide means of ascertaining the signal level at various points in the network.

## *LEVEL INDICATORS*

**The Volume Indicator.** One of the most important components in a broadcast-station network from the point of view of utility is the volume indicator. This instrument permits the control operator to monitor a program continually by accurate visual means. It enables him to make precise adjustments of incoming levels from various microphone or other pickup channels and also provides him with a continuous visual indication of the level that he is relaying, via transmission lines, to the transmitter speech amplifier.

An elementary circuit of a typical volume-level indicator is shown in Fig. 301. The grid of the tube is maintained at the proper negative bias by means of the potentiometer to enable it to function as a linear rectifier. The input circuit may be either resistance-coupled or transformer-coupled to the a-f line being measured. In either case, the input impedance is kept sufficiently high to provide negligible shunting effect across the line.

The plate-circuit galvanometer is calibrated directly in decibels. The a-f input signal voltage energizes the grid of the tube, causing a resultant plate-current flow and deflection of the galvanometer needle. Taps on the secondary of the input transformer permit extending the range of the instrument by reducing the input at the higher levels. Each tap is calibrated directly in decibels to permit proper correction to be made to the galvanometer reading. Some types of instruments employ several scales on the galvanometer dial, permitting direct readings for all input ranges.

A few of the older-type volume indicators utilize bimetallic rectifiers in place of the vacuum tube, but such instruments operate on the same essential principle. The a-f signal voltage is rectified by a bimetallic

rectifier of the copper oxide type. The rectified output flows through a sensitive galvanometer calibrated directly in decibels. Such units are essentially rectifier-type voltmeters, except that the meter scale reads energy levels in decibels instead of in volts.

Volume indicators are sometimes referred to as "power-level indicators," since the output-meter scale is calibrated in actual decibels above zero level. In most broadcast work, the standard reference level of 6 mw in 500 ohms is used as zero-decibel level. Some types of equipment, however, are designed with a zero decibel level of 12.5 mw. When the power level represented by zero decibel is known, the actual power in the circuit at any other decibel level can easily be computed as described in Chap. XIV.

The ideal volume-level indicator is one that draws zero current from the measured circuit and has an indicating needle that instantaneously follows all variations in signal level. In actual practice, such performance

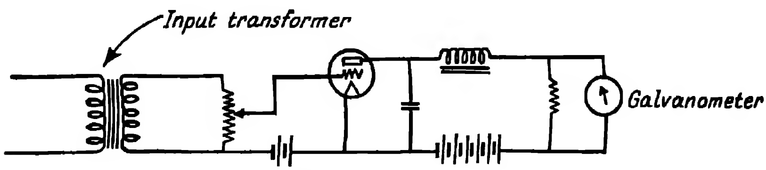



FIG. 301. Electronic volume indicator.

is impossible to attain, because of input circuit impedances and the natural inertia of the galvanometer indicating needle.

All vacuum-tube volume indicators are essentially a form of *electronic*, or *vacuum-tube, voltmeter*. A vacuum tube with its grid negatively biased offers an extremely high input impedance. If it were possible to connect the tube grid-cathode circuit directly across the circuit to be measured, there would be no shunting effect at all, since the negative grid prevents any current from flowing in the grid-cathode circuit. Unfortunately, however, an intervening circuit is necessary between the input and the grid circuit to provide bias voltage to the grid and to permit adjustment of input levels. The input impedance offered by the unit as a whole is therefore primarily determined by the input circuit network.

Many modern types of volume indicators utilize a linear preamplifier before the rectifier tube, an arrangement that provides much higher input impedance. By perfect linear operation of the amplifier, no distortion is introduced into the circuit. For such units high-speed indicating meters have been developed that reach full deflection in approximately 0.15 sec. Since this time unit approaches the time required for the shortest sound pulse occurring in a program, a high degree of accuracy is possible.

The high-speed needle action by itself would result in such violent fluctuations of the pointer that accurate readings would be very difficult

to obtain and constant monitoring would become extremely tiring. This objection is overcome by the use of a slow discharge circuit, which permits the needle to rise very rapidly on peaks but retards the falling off from these peaks. This control is accomplished by a capacitor having a very high resistance load. The rectified signal simultaneously energizes the grid of an amplifier tube and charges this capacitor. The galvanometer is in the plate circuit of this tube. When the signal voltage decreases following a peak, the grid is controlled by the discharging condenser voltage. Since this discharge occurs through a high resistance, the plate current falls off slowly. The result is a "floating" movement of the galvanometer needle.

## *BROADCAST-STUDIO AMPLIFIERS*

The theory of a-f amplifiers has already been discussed in detail in Chap. XII. The amplifiers used in broadcast-studio equipment are identical to receiver a-f amplifiers so far as the principle of operation is concerned. However, the design is altered in a number of respects for broadcast use.

Broadcast-studio amplifiers are generally classified according to their application, such as *preamplifiers*, *program amplifiers*, and *line amplifiers*. All amplifiers used in broadcast work are high-fidelity amplifiers. Modern studio amplifiers have essentially linear frequency response from 30 to 10,000 c and are far superior to amplifiers usually found in receiving equipment.

**Preamplifiers.** Modern microphones have comparatively minute power outputs and require a tremendous amount of amplification. As the number of stages in an amplifier is increased, it becomes exceedingly difficult to secure stable operation because of feedback coupling with its attendant distortion and oscillation tendencies. For this reason, it is unfeasible to obtain the great amount of amplification required in a single high-gain amplifier. Therefore, a number of individual low-gain amplifiers are utilized. It is usually possible to place such amplifiers sufficiently far from each other to avoid harmful coupling. By utilizing low gain per stage, it is possible to design exceedingly efficient and stable amplifiers of three or four stages that have excellent fidelity characteristics.

The main component unit in a studio amplification system is the program amplifier. This amplifier handles the output of all studio channels, such as studio microphones, lines from remote pickups, and electrical transcription apparatus. In order to facilitate monitoring and mixing operations and to provide the proper degree of mixing when several different channels are in operation simultaneously, it is necessary that the incoming levels on each channel be approximately equal. The control operator is then enabled to control relative channel levels by

attenuating the desired channels. Since the output of a microphone is very much lower than that of other types of studio equipment, it is necessary to supply considerable amplification to microphone channels *before* the program amplifier. This amplification is provided by the *preamplifier*.

In addition to performing its function as an amplifier, the preamplifier, when located close to the microphone, as is customary, greatly decreases extraneous noise levels by such usage. If the microphone were coupled directly to the program amplifier, a comparatively long line would have to be used. Despite thorough shielding and grounding, stray fields induce disturbances, or noise voltages, in such lines. Because of the extremely low power levels delivered by the microphone, even minute disturbances in the line will be of a magnitude comparable to the program level. Such disturbances will therefore be subsequently amplified by the program amplifier to the same approximate extent as the program. The result is very noisy reproduction and low signal-to noise ratio.

By locating a preamplifier close to the microphone, the power level in the line to the program amplifier will be correspondingly greater. Noise disturbances picked up by the line will then be of much smaller magnitude than the program level, and the resultant signal-to-noise ratio will be greater.

Preamplifiers are designed with input impedances of 50 to 200 ohms, depending upon the type of microphone to be matched. The output circuit is customarily designed to work into a 500-ohm line.

Because of the extremely low input power level, all input leads are as thoroughly shielded as possible against magnetic and electrostatic fields All contact connections, such as switches, plugs, jacks, and sockets, must be firm and absolute. All such connections must be periodically inspected and cleaned, since even minute variations in contact resistance introduce serious noise. Input transformers in preamplifiers are a common source of noise owing to the pickup of stray a c fields. These transformers are therefore unusually well shielded with heavy Permalloy magnetic shields. In addition, such transformers are center-tapped, and the tap is grounded in circuits that permit such arrangement. Noise voltages induced in the line between microphone and preamplifier will be equal in both wires of the line if the line is kept properly balanced. The noise currents caused by these voltages will therefore flow in opposite directions through the transformer primary winding and will be balanced out to ground through the center tap.

Often r-f charges from stray fields are picked up by the input circuit of a preamplifier. Under certain conditions, the first tube of the amplifier will act as a detector at such frequencies, producing a rectified signal voltage and causing objectionable noise. This effect is overcome by the use of special input tubes that are capable of linear operation over a

much greater range than that required for the proper transmission of the program.

**Program Amplifier.** The function of the program amplifier is to increase the level of the combined studio channels to the proper value for line transmission to the transmitter speech amplifier. This level is customarily limited by the telephone companies, from whom the lines are leased, to approximately 0 db with permissible peak levels of short duration not to exceed + 2 db. The gain of the program amplifier is controlled by the operator through a master control. In some types of circuits, this control functions actually to vary the gain of the amplifier. However, the approved types of program amplifiers utilize an attenuation network in the output circuit with variable taps. The amplifier is thus permitted to function at a fixed gain, which is determined by the point of best operating efficiency and fidelity.

In addition to providing program levels for the line, the program amplifier must also provide much higher levels to operate a loudspeaker utilized by the control operator for monitoring purposes. Although a separate amplifier is often used for this purpose, both functions are customarily combined in a single amplifier This amplifier provides sufficient gain to operate the speaker. An attenuator network reduces this level to the proper amount necessary to feed the line

In addition to guarding the maximum line levels, the control operator must see that the modulation capabilities of the transmitter are not exceeded. The level going to the line must never be allowed to exceed that value which causes the transmitter to overmodulate. On the other hand, the level must never be allowed to fall below the value at which the over-all circuit noise level exceeds that of the program. Normally, these functions are performed by the control operator by manual manipu lation of the master gain control. In effect, he "compresses" the volume range of the program whenever necessary to keep it within the necessary limits.

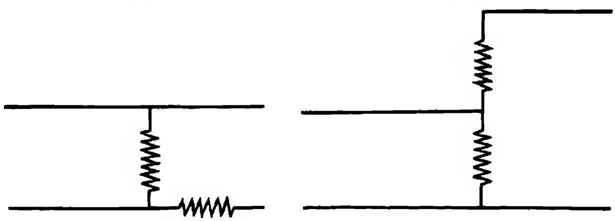
Modern program amplifiers automatically perform the function of *volume compression*, thus relieving the operator of this duty. The gain of such amplifiers can be adjusted in the normal manner, but the output can be limited to a certain level which is predetermined and manually adjusted. Although the amplifier functions over the normal wide range, all output levels in excess of the predetermined value are reduced to this value. This critical value is determined by the power level into the line that will permit 100 per cent modulation of the transmitter.

Volume compression is obtained by feeding a portion of the output voltage into a rectifier. The d-c output of the rectifier is utilized to increase the negative bias on one of the amplifier tubes, thereby decreasing the gain. An adjustable attenuator circuit between the amplifier output and the rectifier input circuits determines the level at which the rectifier

begins to operate and permits adjustment of the compression circuit. By means of this adjustable attenuator control, the unit can be adjusted so that compression will start above any desired level.

## *ATTENUATORS*

Attenuators are networks of resistors that find wide application in broadcast work and are used whenever it is desired to introduce a loss into a circuit, as the name implies. In addition, they are also used as a means of matching circuits having unequal impedances. In such cases, they are designed to introduce a minimum amount of loss into the circuit. Since attenuator networks are composed of resistance only, the inductive



(a) (b)

FIG. 302. (a) L-pad. (b) Equivalent circuit for (a).

and capacitive effects are negligible and the frequency response characteristics of a line or circuit are not altered by their insertion. Where line balance is important, balanced attenuators are available. which have symmetrical proportions that do not disturb line-balance conditions. Where line balance is not important, simpler attenuation networks are available that have characteristics equivalent to the similar type of balanced network.

**Attenuator Networks.** Attenuator networks, more often called **attenuator pads,** or **line pads,** are of three general types, namely, the L pad, the T pad, and the $\pi$ pad. The latter type is often called a "ladder" attenuator because the circuit drawing resembles a ladder when several such pads are combined. The simplest form of attenuator pad, aside from a single resistor, is the L pad, shown in Fig. 302(a). This, it will be seen, is simply a circuit containing a resistor in series with one wire of the line and another resistor shunted across both wires, as shown in Fig. 302(b). An L pad can be used only between circuits of unequal impedances, since it is apparent that one of the pad resistors will always be in series with one of the line impedances but not with the other.

The attenuator pads most widely used are of the T and $\pi$ types, both of which are available as either balanced or unbalanced networks. The balanced T pad is usually called an "H pad" because of the configuration.

The values of the component resistors of an attenuator pad depend upon the input and output impedances of the circuits to which the pad is coupled and upon the inherent loss desired in the pad. Pads can be designed to work between any two impedances, whether equal or unequal. For any given input-to-output (or vice versa) impedance ratio, however, there is a certain minimum loss inherent in the pad. Conversely, for a pad of given loss, there is a certain maximum impedance ratio that may not be exceeded. Thus, the maximum impedance ratio that a 10-db pad may have is 3.018, that is, an output impedance 3.018 times input

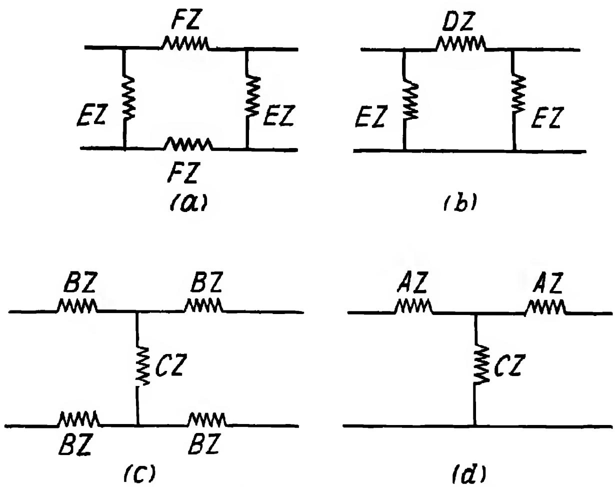



FIG. 303 Equivalent balanced and unbalanced pads for use when input and output impedances are equal. (a) Balanced $\pi$ pad. (b) Unbalanced $\pi$ pad (c) Balanced T (or H) pad (d) Unbalanced T pad

impedance, or vice versa. If any attempt is made to increase the impedance ratio, the loss will become greater than 10 db. Stated in another way, if the impedance ratio of a pad is designed to be 3.018, the loss introduced by the pad must be at least 10 db Pads for this impedance ratio can be designed having any loss *greater* than 10 db. If, however, any attempt is made to decrease the loss to a value less than 10 db, the impedance ratio will fall below the desired value of 3.018.

Formulas have been derived for each resistor in an attenuator pad. These formulas give the resistance value in ohms in terms of the input and output impedances combined with a loss constant, or multiplier. The loss constants for T and $\pi$ pads of both the balanced and unbalanced types are tabulated in Table VI of the Appendix for the most frequently required losses. The maximum possible impedance ratio for each loss value is also shown.

Attenuator networks for balanced and unbalanced T and $\pi$ pads are

shown in Fig. 303. The formulas shown for the resistor values are for the simple case where input and output impedances are equal. The common value of input and output impedance is represented by the letter $Z$, and the remaining letters refer to the loss constants listed in Appendix Table VI.

The design of attenuator networks to be used between unequal impedances involves considerably more calculation. The formulas for the component resistor values in the balanced and unbalanced T and $\pi$

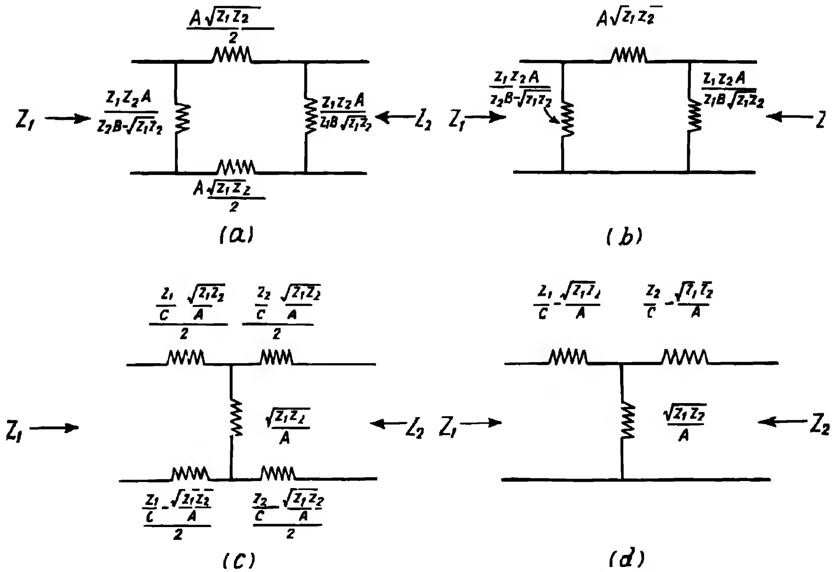



FIG. 304 Equivalent balanced and unbalanced pads to be used between unequal impedances (a) Balanced $\pi$ pad (b) Unbalanced $\pi$ pad (c) Balanced T (or H) pad (d) Unbalanced T pad

pads are shown in Fig. 304. In these formulas, $Z_1$ represents input impedance, $Z_2$ represents output impedance, and the remaining letters refer to the loss constants of Table VI

**Problem.** It is desired to design an H-type (balanced-T) pad to work between equal impedances of 500 ohms that will have a loss of 15 db.

**Solution.** The pad desired is of the type shown in Fig 303(c) The loss constants $B$ and $C$ for a 15-db pad are found from Table VI to be

$$B = 0.34905, \tag{1}$$

$$C = 0.3672. \tag{2}$$

Then,

$$\text{series element} = BZ = 0.34905(500) = 174.525 \text{ ohms}, \tag{3}$$

$$\text{shunt element} = CZ = 0.3672(500) = 183.6 \text{ ohms}. \tag{4}$$

The foregoing work can be checked by drawing the equivalent series-parallel circuit of the network as it appears from either end and including the line impedance of the opposite termination. This is shown in Fig. 305. The upper series leg of the middle parallel circuit is

$$BZ + Z_0 + BZ = 175 + 500 + 175 = 840 \text{ ohms,} \tag{5}$$

where $Z_0$ = output impedance = 500 ohms;

$BZ$ = 175 ohms = approximate value taken from Eq. (3).

The net resistance of the middle parallel circuit is

$$\frac{840 \cdot 184}{840 + 184} = \frac{154{,}560}{1{,}024} = 150 \text{ ohms,} \tag{6}$$

where $CZ$ = 184 ohms = approximate value taken from Eq. (4).

The resistance of the entire series-parallel combination then becomes

$$BZ + 150 + BZ = 175 + 150 + 175 = 500 \text{ ohms.} \tag{7}$$

Since 500 ohms is the desired input impedance, the network checks for proper termination. The network could be checked for termination by looking

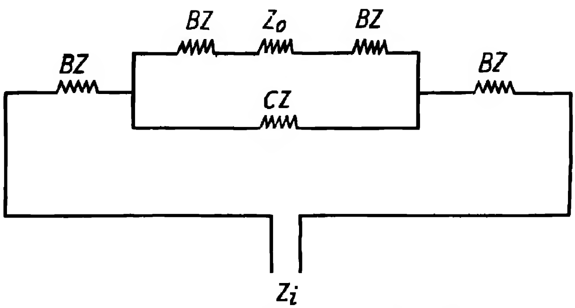



FIG. 305. Equivalent series-parallel network of a balanced T pad.

into the output end in a similar manner. The only change in the equivalent circuit, then, would be substitution of $Z_i$ for $Z_0$. Since the input and output impedances are equal, the foregoing proof suffices for both terminations. It is seldom necessary to check such a circuit from both terminations, even when input and output impedances are unequal. If the network checks from one end, it must, of necessity, check from the other end.

The network can be checked for power loss by assuming an input power of some convenient figure applied to either end. The current distribution through the equivalent network is then calculated and the power expended in the output impedance determined. The ratio between input and output power is then computed and converted into decibels. Thus, assuming a power input of 50 w in the pad of the foregoing problem, the current flowing into the circuit is

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{50}{500}} = 0.316 \text{ amp,} \tag{8}$$

where $R$ = 500 ohms = total circuit resistance, including output impedance.

The voltage across the middle parallel circuit of Fig. 305 is then found to be

$$E = IR = 0.316(150) = 47.4 \text{ volts.} \tag{9}$$

The current in the upper leg of the middle parallel circuit is

$$I = \frac{E}{R} = \frac{47.4}{840} = 0.056 \text{ amp.} \tag{10}$$

The power expended in the output impedance is

$$P = I^2Z = (0.056)^2 500 = 1.57 \text{ w.} \tag{11}$$

The power ratio between input and output powers is thus found to be

$$\text{Power ratio} = \frac{P_i}{P_o} = \frac{50}{1.57} = 31.6. \tag{12}$$

From the fundamental formula for the decibel (see Chap. XIV)

$$N_{db} = 10 \log \frac{P_i}{P_o}. \tag{13}$$

Substituting Eq. (12) in Eq. (13),

$$N_{db} = 10 \log 31.6 = 10(1.499) = 14.99 \text{ db.} \tag{14}$$

The attenuator pad therefore checks with an insertion loss of approximate 15 db.

Attenuator pads have three major applications in broadcast work. The most common use is for the purpose of introducing a loss in the circuit. This application is illustrated in the program amplifier discussed in the previous section, where an attenuator pad is utilized to decrease the power level at the monitoring loudspeaker output to the proper level for transferring to the transmitter line. This is an example of purposely inserting pad loss in a circuit.

Pads are often used when patching circuit connections in preparation for remote line pickups, as, for example, when it is desired to couple a 500-ohm line to a 200 ohm amplifier input. In such cases, the pad is designed for minimum loss consistent with the impedance ratio desired. This loss is usually easily recovered by increasing the gain of the succeeding amplifier. In this application, as in the foregoing one, attenuator pads have the advantage of economy and of not affecting the frequency response characteristics of the circuit. Since all the pad elements are resistances, the insertion loss is a constant at all frequencies.

Attenuator pads also find considerable application as circuit-isolation devices. Since transmission lines and transformers are true impedances, they vary with the frequency. If an amplifier output transformer is coupled directly to a transmission-line load, the transmission line and transformer impedances will vary considerably at different frequencies. This variation causes a corresponding variation in the reflected plate load in the final amplifier stage, with a resultant variation in amplifier-frequency response. The insertion of a balanced attenuator pad between

transformer and line will greatly minimize this impedance fluctuation. This can be seen by an inspection of the H-pad equivalent circuit of Fig. 305. If $Z_t$ represents the transformer terminal impedance and $Z_0$, the impedance of the line, it is apparent that even a direct short circuit of $Z_0$ would not reduce the impedance presented to the transformer to zero. Similarly, if $Z_0$ were open-circuited, the load presented to $Z_t$ would still be considerable. If the transformer were connected directly across the line without the pad, it is obvious that under the above conditions the impedance across the transformer would vary from zero to infinity.

It can also be seen by inspection of the diagram that the larger the resistance value of the pad elements, the smaller the proportion of total impedance variation caused by a given fluctuation of load impedance. In other words, the higher the pad loss, the greater its effectiveness as an isolation unit. Actually, the total impedance variation across the transformer terminals using the 15-db pad of the foregoing problem would be approximately 12 per cent under the extreme conditions of a variation from short circuit to open circuit of the line. By using a 40- or 50-db pad the net impedance variation under the same conditions could be reduced to approximately 0.01 per cent. Since, in actual practice, line impedance variations cover a much more limited range, relatively low loss pads are used with great effectiveness for isolation purposes.

## *MIXER CONTROL UNITS*

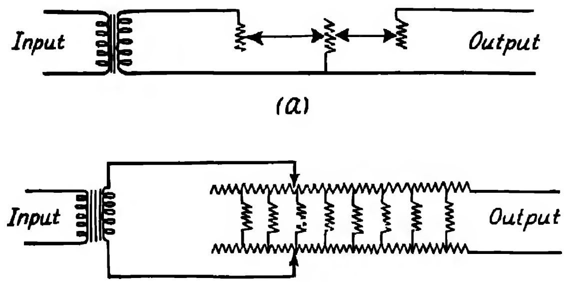
**Faders.** One of the problems confronting the broadcast technician is that of feeding several individual channels into a common program amplifier in a manner that permits individual control of the channel levels. Such control enables the operator to fade individual channels up or down. Thus, a musical program can be faded down to permit announcements, advertising talks, or noise effects to be superimposed upon the program. The variable network in each channel, by means of which such adjustments are made, is called a **fader.**

Faders are variable attenuator pads. In addition to providing a variable loss by means of tapped or continuously variable resistor elements, faders also fulfill the function of maintaining the same input and output impedances, regardless of the attenuation setting. All faders in common use are variations of the conventional T and H attenuation pads. Most commercial faders permit a variation in loss of approximately 40 to 50 db in steps of 2 db or less.

Even when single-channel input to the program amplifier is utilized, faders are used for volume control in preference to variation of the amplifier gain. It is customary to adjust the program amplifier gain to the maximum value that provides the desired frequency-response characteristic as ascertained by an over all frequency run. Once this

adjustment is obtained, the amplifier gain is unchanged to prevent possible variation of frequency response. Since faders are resistance networks that have negligible inductance and capacitance, the over-all frequency response of a system is unaffected by their insertion or variation.

The faders most widely used in broadcast work are the unbalanced T type and the H or ladder type The circuit diagram of a typical T-type fader is shown in Fig. 306(a) The variable tap switches are ganged so that any increase in series resistance is accompanied by the proper decrease in shunt resistance to maintain the same input and output impedances. The values of the series and shunt elements are determined in the same manner as those of a conventional T pad. The maximum value of the series resistors and the minimum value of the shunt resistor




(b)

FIG. 306 (a) T type fader (b) Ladder type fader.

are determined by designing a fixed T pad having the desired maximum loss and proper terminal impedances. Another set of values for a T pad are computed to obtain a pad of minimum loss at the given impedance ratio These values determine the taps on the first resistors for minimum loss setting of the attenuator and correspond to the maximum permissible value of shunt resistance elements and minimum permissible value of the series resistance elements. The values for the intermediate taps are computed in the same manner Each set of tap settings must be calculated separately for the desired loss, maintaining the same terminal impedances. T-type faders are used in channels where line balance is not of paramount importance, since the network is essentially unbalanced because all the series resistance is in one side of the line.

The faders used for balanced lines are of the ladder type shown in Fig 306(b) It will be seen that the ladder fader is essentially a number of H-type attenuators in series. Loss variation is obtained by adjustment of the input circuit taps. The tap controls are ganged together to provide a balanced circuit at all settings. Moving the tap pointers from left to right in Fig. 306(b) decreases the attenuator loss. The position shown in the diagram is for maximum loss. As the tap pointers are moved to the

right, the series resistance in both legs is decreased. Simultaneously, the shunt resistance is increased because the series elements to the left of the taps become a part of the shunt circuit, which automatically preserves the original terminal impedances at all positions of the tap switches.

Ladder-type faders are designed in the same manner as H-type attenuator pads. Owing to the greater number of resistor elements, the calculations become correspondingly more involved than those for the T-type fader.

**Mixers.** When a number of faders for individual studio channels are simultaneously fed into a common program amplifier, the resulting network is called a **mixer.** A number of typical T-type and H-type mixers are shown in Fig. 307. Most modern broadcast stations use series or series-parallel coupling of the individual faders to the program-amplifier input transformer. Parallel coupling is often used when balanced faders are available in all channels. When unbalanced faders are used, the percentage of total output impedance shift for a given variation of a single-fader output impedance becomes undesirable with parallel coupling.

A series mixer circuit of four channels utilizing T-type faders is shown in Fig. 307(a). Since the fader outputs are connected in series, the output impedance of each fader should be one fourth the required amplifier input impedance. Thus, for an amplifier of 500 ohms input, the output terminal impedances for each of the four faders of Fig. 307(a) should be 125 ohms.

Figure 307(b) illustrates a four-channel T-type mixer utilizing series-parallel connection. The output impedance of each fader in this case should be the same as the input impedance of the amplifier. Thus, for a 500-ohm amplifier input, the output impedance of each fader should be 500 ohms. Each pair of parallel connected faders will have an output impedance of 250 ohms. Both of the 250-ohm impedances in series will result in a total output impedance for the network of 500 ohms.

A series-connected four-channel ladder type mixer is shown in Fig. 307(c). The output impedance is derived in the same manner as that of the T-type mixer of Fig. 307(a). Although the ladder-type mixer is, in general, more efficient than the other types, for economical reasons, it is not so often found in practice.

Most mixing is done at high program levels *after* preamplification. Generally speaking, the higher the level at which mixing is accomplished, the better. Although all contacts in mixer circuits are of the "wiping" type, some noise does originate in these circuits. At high program levels, the noise voltage is inconsequential compared with the signal, and a certain amount of noise can be tolerated.

When, as is often true, low-level mixing is employed for practical reasons, any noise voltages resulting from disturbances in the mixer circuit are of a level comparable to the signal. They therefore receive

approximately equal amplification with the signal in the succeeding amplifiers and assume troublesome proportions in the output. Consequently, special care must be taken with faders used in low level mixing circuits to ensure freedom from dust, dirt, and oxidation of contact surfaces.

In order to perform properly the operation of mixing the control

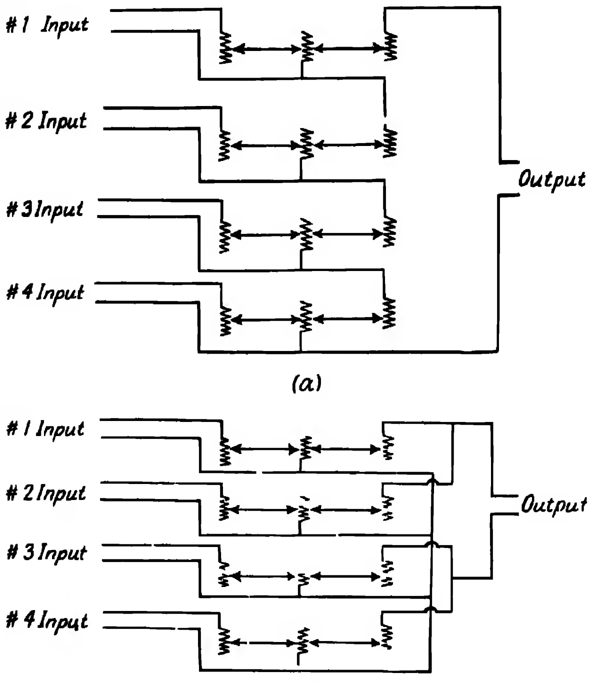


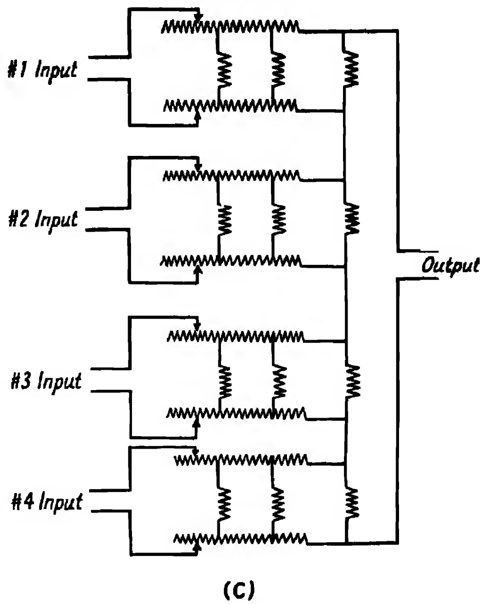

Fig. 307. (a) Four channel T type series mixer. (b) Four channel T type series parallel mixer.

operator is provided with a volume indicator that can be bridged across the individual fader outputs. Although a separate volume indicator for each channel is desirable the cost is usually prohibitive, and a single indicator is used with a switching arrangement.

## EQUALIZERS

One of the objectives of radio broadcasting is to transmit a studio program with the same lifelike quality with which it is heard in the place where it originates. In order to achieve such high fidelity reproduction, all the equipment utilized in relaying the program to the transmitter

must have the same high-fidelity characteristics as the transmitter-modulator and speech-amplifier circuits. In modern broadcasting, essentially linear frequency response from 30 to 10,000 c is required of all program equipment. Since modern studio amplifiers and microphones are designed to meet these specifications. the only remaining components in a program network in which this element is a variable, are the studio links, that is, the transmission lines used to connect studio to transmitter, remote pickup to studio, and so on.




307(c). Four-channel ladder-type series mixer.

The frequency response of a transmission line is a function of its distributed inductance and capacitance. The response of a line therefore varies with its length. Since a great many different circuits are set up in a busy broadcast station throughout the course of a day, it is evident that the frequency-response characteristics of the various lines must often be checked. Checking is accomplished by means of *frequency runs* on the lines. From the frequency run, the actual frequency characteristics of a line are obtained. When the response characteristic is not essentially linear within approximately plus or minus 2 db over the desired range, it is desirable to alter the characteristic of the line. This is accomplished by means of *equalization networks*, or *equalizers*, which are inserted across the line.

The equalizers used in broadcast work consist of an inductance and capacitance in parallel, connected in series with a variable resistor across the line, as shown in Fig. 308. Most transmission lines are lacking in h-f response because of the shunting effect of the capacity between the lines. Since capacitive reactance varies inversely with the frequency, the reactance becomes very low at the high frequencies and effectively causes a partial short circuit at these frequencies. The longer the line, the greater the capacitance and the lower the capacitive reactance become. Hence, the h-f attenuation increases with line length.

It is impossible to decrease the line attenuation at high frequencies.

However, it *is* possible to increase the attenuation at low frequencies with the circuit shown in Fig. 308. The $L$ and $C$ constants of the circuit are chosen to make the circuit resonant at a fairly high frequency, customarily 5,000 c. At this frequency, the parallel resonant circuit offers almost infinite impedance and has negligible shunting effect across the line. With the proper choice of circuit constants, the impedance falls off relatively slowly at frequencies higher than resonance. At frequencies lower than resonance, the impedance drops quite rapidly. The shunting effect at low frequencies is quite pronounced and varies inversely with the frequency. The extent to which the low frequencies are by-passed is determined by the setting of the series-variable resistor. By judicious

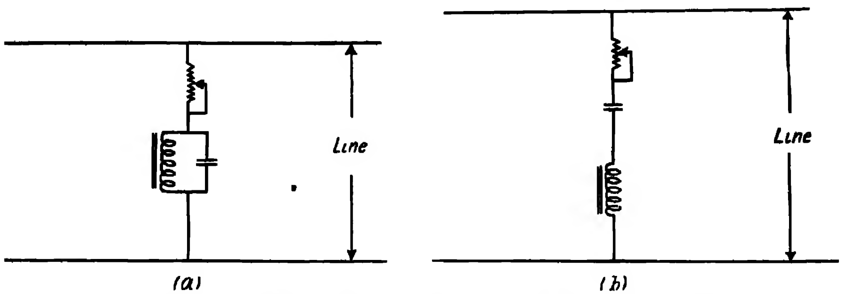



FIG. 308. (a) Parallel-circuit equalizer (b) Series-circuit equalizer.

adjustment of this resistor, essentially linear response can be obtained by making the l-f loss through the equalizer equal to the h-f loss in the line.

A similar effect may also be obtained by the series circuit equalizer shown in Fig. 308(b). In this case, the $L$ and $C$ components are chosen to produce series resonance at a low frequency. At this frequency the circuit offers minimum impedance (equal to the resistance), and the low frequencies are effectively by passed. As the frequency is increased above resonance, the impedance of the series circuit increases. At the very high frequencies, the shunting effect is negligible.

## THE MODULATION MONITOR

All broadcast stations are required by the FCC to have in operation a modulation monitor of a type approved by the Commission. Such monitors provide a continuous indication of the percentage modulation and are usually used in conjunction with controls that permit the monitoring operator to vary the modulating voltage.

Modulation monitors operate on a small r-f voltage obtained from the transmitter output. A typical commercial modulation-monitor circuit is shown in Fig. 309. Essentially, the monitor circuit consists of a diode rectifier, a suitable r-f filter circuit, and a metering circuit.

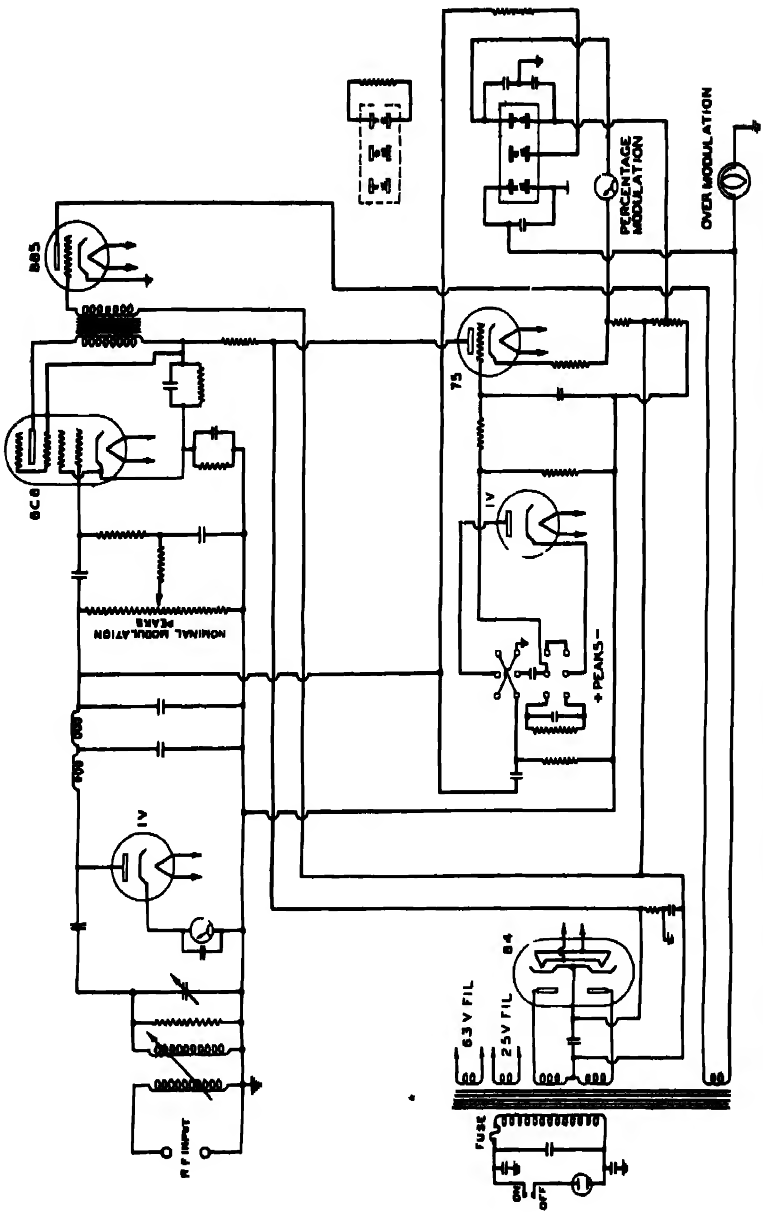



FIG. 309. Circuit of a typical modulation monitor. *(Courtesy of General Radio Company)*

The metering circuit consists of an audio rectifier and a vacuum-tube voltmeter.

The r-f input to the modulation monitor is controlled by a series-variable attenuator or by variation of the input transformer coupling and is applied to a diode rectifier (detector). The r-f component in the output of the detector is removed by the r-f filter. The pulsating d-c output therefore contains only the a-f (modulation) component. The a-f voltage developed across the diode load resistor by this component is rectified by the metering-circuit diode rectifier and is used to charge a capacitor. The capacitor has a very high resistance load; consequently, although it charges very rapidly, it discharges at a relatively slow rate. The voltage across this capacitor controls the grid of a triode connected as an electronic voltmeter. The plate current of the triode is proportional to the degree of modulation, and a microammeter in the plate circuit is calibrated to read directly in percentage modulation.

High-speed action of the meter needle is necessary in order to follow the peaks of modulation. However, as in the case of the volume indicator previously mentioned, this action in itself would result in violent fluctuations of the needle and make accurate readings very difficult. For this reason the high-resistance load is placed across the capacitor in the metering circuit. The slow discharge circuit allows the needle to rise very rapidly on peaks, but the falling off from these peaks is markedly retarded, thus greatly increasing the accuracy of readings.

Many commercial modulation monitors utilize *peak-flash indicators*, which operate to light a lamp or ring an alarm bell through a relay circuit when predetermined peak levels are exceeded. Such circuits consist of an amplifier driving a grid-controlled gas triode. The monitor in Fig. 309 utilizes a type 6C8 amplifier driving an 885 gaseous triode. The gaseous triode is normally biased beyond cutoff. The audio voltage across the load resistor of the monitor first diode (r-f diode) is used to drive the amplifier. Whenever the peak value of the diode output audio component exceeds the amplifier grid bias the grid becomes positive, and plate current flows. The plate voltage drop developed across the transformer primary is stepped up and applied to the 885 grid, overcoming the cutoff bias. The gaseous tube triggers, or discharges, and current flows in its plate circuit. This current operates a lamp, which provides visual indication of excessive modulation peaks. A relay may be connected in the circuit in place of the lamp and the contacts utilized in a number of ways to operate auxiliary indicating devices.

Sufficient r-f pickup to operate the modulation monitor can be obtained by placing a small coil of a few turns close to the antenna lead or to the final amplifier tank circuit. By means of a low impedance transmission line, the coil is coupled to the monitor input circuit. Sufficient coupling must be provided so that the carrier meter in the monitor input diode

circuit can be set to 100 after the signal has been tuned to resonance. Some margin must be allowed in this setting to provide for transmitter-output variations.

The tuned circuit in the monitor is damped by two resistors in parallel, which considerably broadens the resonance peak of the circuit and prevents possible cutting off of side bands. If one of the resistors is removed, somewhat less power will be required to operate the instrument, but side-band clipping will be increased. If more than half the available

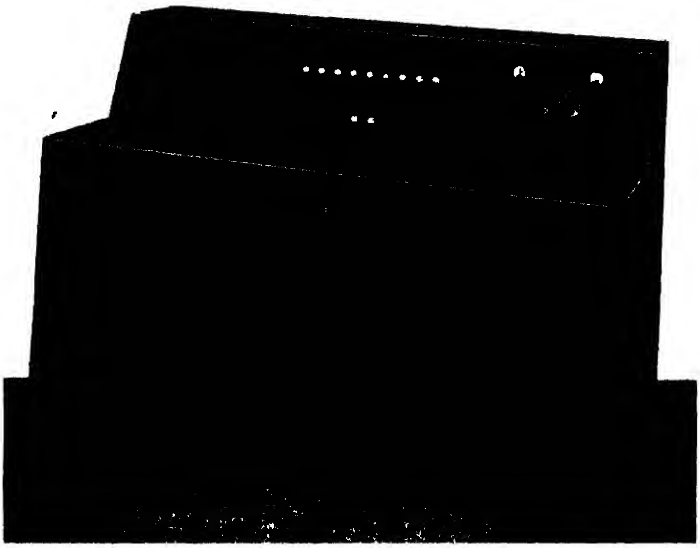


FIG. 310. Control console containing transmitter controls and audio modulation, and frequency-monitoring equipment (*Courtesy of Radio Engineering Laboratories, Inc.*)

tuning capacitance is in the circuit, however, the cutoff action will not be extreme.

## *QUESTIONS AND PROBLEMS**

**1.** What is the purpose of H- or T-pad attenuators?

**2.** Draw a simple diagram showing four mixers connected in series parallel, and using compensating resistors and feeding a balanced load with proper matching.

**3.** Why are preamplifiers sometimes used ahead of mixing systems?

**4.** What is the purpose of a variable attenuator in a speech-input system?

**5.** Why is a high-level amplifier that is feeding a program transmission line generally isolated from the line by means of a pad?

* These questions and problems are taken from the "F.C.C. Study Guide for Commercial Radio Operator Examinations."

**6.** Why are program circuits that are using telephone lines usually fed in at a level of about 12 mw?

**7.** What is the purpose of a line equalizer?

**8.** Draw a diagram of an equalizer circuit most commonly used for equalizing wire-line circuits.

**9.** What methods are employed to avoid switching clicks in switching operations of mixing circuits?

**10.** Why is it generally unnecessary to equalize a short wire line program circuit?

# Chapter XIX

# TELEVISION AND FACSIMILE

The art of transmitting pictures or images by electrical means is not new but dates back to the experiments of the Chinese researcher Nipkow in the late nineteenth century. The rotating scanning disk with spirally disposed apertures, which is the basis of most mechanical television scanning systems, is credited to Nipkow. The number of systems that have been developed for the transmission by radio of pictures (radio facsimile) and of images (television) is legion. Until the advent of electronic scanning, however, it may safely be said that the results were far from completely satisfactory, and there are a number of problems still to be solved. Nevertheless, tremendous progress has been made in the last decade in both the theoretical and the practical approach to the task.

## *TELEVISION*

Basically, television consists of the transmission of an image (which may or may not be moving) so that the reception is to all intents and purposes instantaneous so far as the observer is concerned. In other words, the element of *time* must be considered. This is one important respect in which television differs from other means of transmitting intelligence. Thus, in the transmission of sounds by radiotelephone, a single tone of specified duration is converted into an a-f electric current. By the process of modulation, radio transmission, and so on, this a-f electric current is reproduced at the radio receiver where it is converted into a sound consisting of a tone with the identical frequency and duration of the original sound. In television, an entire scene cannot be transmitted at once, as can a single tone or combination of tones It is necessary to break down the scene into a multitude of smaller scenes, actually small spots, or dots. Each dot must be transmitted separately and received separately. In order that the viewer may observe the scene as a whole, all the dots composing a scene must be transmitted within a certain time minimum. The limiting time interval is a physiological factor and is a function of that peculiar inertia of the human eye referred to as **persistence of vision.** When it is realized that in order to reproduce an image 2 in. square with good resolution, it is necessary to transmit approximately 150,000 dots in a small fraction of a second, the importance of the time element can be appreciated. Furthermore, in order to produce the illusion

of motion, it is necessary that a given image be completely transmitted at least sixteen times per second. In present-day electronic television systems, 30 **frames**, as they are called, are transmitted per second because of technical limitations to be discussed in a later part of this chapter. This means that to reproduce a moving image only 2 in. square, 150,000 dots must be transmitted 30 times each second—a total of 4,500,000 dots per second.

The phenomenon of persistence of vision is due to a peculiar characteristic of the human eye. The retina of the eye will retain an image for a fraction of a second *after* the scene being viewed has been removed. This characteristic is utilized in motion pictures to produce the illusion of motion. If a number of pictures exactly alike are presented to an observer rapidly enough, the observer thinks he has seen but one picture. The first picture is retained by the eye until the second one appears; the second one is retained until the third appears, and so on. It has been found that the pictures must be presented with a frequency of at least 16 pictures (frames) per second to produce this illusion.

In motion pictures, a series of progressive photographs are made depicting the various stages of a movement (of a person, for example). When these pictures are presented to an observer in proper order at a speed in excess of 16 pictures per second, he thinks he sees a single continuously moving picture.

Television would be impossible were it not for the persistence of vision characteristic of the human eye, since because of this characteristic a scene can be broken up into a number of small spots. If the *last* spot of the 150,000 in the image previously mentioned is transmitted within $\frac{1}{16}$ sec of the *first* spot the first spot will still be retained by the eye as well as all the intervening spots. Consequently, although the observer is actually viewing a succession of minute spots, the result appears to him as an entire scene. By transmitting a minimum of 16 such scenes, or frames, per second, the illusion of a moving scene is created.

The number of spots into which a scene must be resolved for satisfactory television is also dependent upon a characteristic of the human eye. If a newspaper photograph is closely observed, it will be seen that it is composed of a number of small dots of ink of varying degrees of darkness. If a magazine photograph is closely observed, it will be seen that it is likewise composed of myriads of small dots. In the magazine, however, the dots are smaller and closer together. When compared with the newspaper picture, the magazine picture will be seen to be much clearer. The difference is a matter of *resolution*. The magazine picture is said to have good resolution; and the newspaper picture is said to have relatively poor resolution.

If a high-grade photograph is closely scanned by an observer, probably no dots at all will be observed. The surface of the photograph will appear

to be continuous. Nevertheless, if this photograph were inspected under a powerful lens, it would be seen to be composed of dots also. The dots would be microscopically small and almost infinitely close together. The limit beyond which the human eye is incapable of distinguishing such dots, as the dots and the separation between them are made smaller, is called the **limit of resolution of the eye.** Because of the fact that there *is* such a limit, it is possible to obtain high-fidelity reproduction by means of dot transmission. The number of dots that it is necessary to transmit to produce excellent reproduction (in this respect) of a scene of given size depends upon the average finite resolving power of the human eye.

**The Television Transmitter.** The heart of a television transmitting system is the scanning unit. There are two methods of electronic scanning used with success in this country. The first utilizes an *image dissector* developed by P. T. Farnsworth as the scanning unit, and the second utilizes the *iconoscope* developed by V. K. Zworykin The system employing the iconoscope will be described in this chapter.

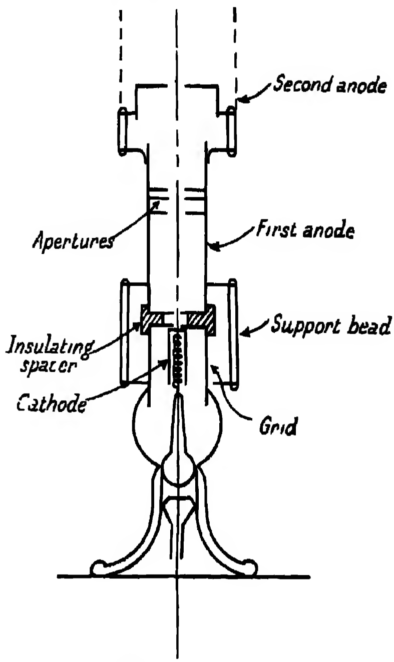



FIG. 311. Elementary iconoscope electron gun.

The iconoscope depends for its operation upon an *electron gun* similar to that used in the conventional cathode-ray tube The details of an elementary iconoscope electron gun are shown in Fig. 311. By means of two positively charged electrodes disposed as shown in the illustration, the emission of a hot cathode is concentrated into a narrow beam of electrons. When the electron beam is passed between two horizontal plates having a difference of potential between them, the beam is bent because of the attraction of the electrons toward the more positively charged plate. Since the plates are horizontally mounted, the deflection of the electron beam will be in a vertical direction, that is, toward either the upper or the lower plate. Similarly, horizontal deflection of the electron beam is accomplished by passing it between a pair of vertical plates. The beam can then be directed, very much like a stream of water from a hose nozzle, by the disposition of the voltages on the horizontal and vertical deflecting plates.

In some types of iconoscopes, deflection of the electron beam is accomplished by the action of magnetic fields upon the beam. The two components of deflection at right angles to each other can be obtained by two coils so mounted that their magnetic fields are perpendicular to each

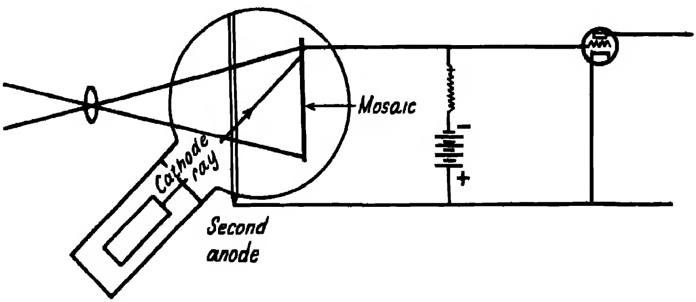



FIG. 312. Elementary iconoscope coupling circuit

other. Usually, however, a single coil of special design is utilized to achieve this effect.

An elementary iconoscope is shown in Fig. 312. The electron gun is in the long cylindrical neck of the tube. A photosensitive plate, or screen, is so disposed in the enlarged portion of the tube envelope that the electron beam can be directed to every part of its area. The optical image of the scene to be televised is focused upon the plate, and the plate is placed at an angle so that the electron gun does not interfere with the optical path of the scene. Upon the illuminated side of the plate is a mosaic surface of minute silver globules that are photosensitized and insulated from one another. The plate itself is composed of an insulating material, usually a very thin sheet of mica. The other side of the plate is coated with a metal film to form a conducting electrode known as the **signal plate.**


FIG. 313. Scanning pattern

Each globule of the mosaic forms a small capacitor with respect to the signal plate. When light from a scene falls upon a globule, electrons are lost through photoelectric emission. The globule therefore becomes positively charged with respect to the signal plate. The magnitude of the charge is proportional to the light intensity.

By means of saw-toothed voltages applied to the horizontal and vertical deflecting fields of the electron gun, the electron beam is caused to scan the mosaic plate in a number of successive lines from left to right,

as shown in Fig. 313. The generation of saw-toothed waves and the mechanics of scanning are enlarged upon in a following section.

When the electron beam strikes a globule that has been positively charged through photoemission, the negative charges that have been lost are replaced by the electrons in the beam. The beam, in effect, therefore, discharges the globule. At the instant of discharge there is a flow of current through the resistance $R$ proportional to the positive charge accumulated on the globule. The voltage drop across $R$ is then proportional to the globule charge and hence proportional to the intensity of the illumination upon this particular globule.

It will be seen, therefore, that the voltage applied to the input circuit of the amplifier tube continually varies as the spot of the electron beam passes from globule to globule. The extent of this variation depends upon the relative intensities of illumination upon adjacent spots of the mosaic. The *amplitude* of the a-c component in the amplifier plate circuit at any instant is therefore a function of the intensity of the illumination upon the globule being scanned at that instant. The *frequency* of the plate circuit a-c component depends upon the rapidity with which globule illumination intensities change during scanning. Since there are countless gradations in contrast in the average scene, it is apparent that the signal being amplified will contain frequency components that vary from low values to extremely high values.

The extreme case requiring the highest frequency would be represented by an image of a checkerboard pattern of alternate black and white squares with the side of each square equal to the width of the scanning line. In scanning such a scene, the iconoscope output would be alternately large and small as the beam spot passed from a white to a black square. One cycle of output current would be represented by the scanning of two successive squares, one white and one black. Since the frequency of an alternating current is expressed in cycles *per second*, the frequency that the iconoscope amplifier would be required to transmit in order to amplify the checkerboard-image signal would equal the number of pairs of black and white squares scanned *per second*. Expressed mathematically,

$$\text{required frequency band} = \frac{a^2bn}{2}. \tag{1}$$

where $a$ = frame height in scanning lines;
$b$ = aspect ratio (ratio of frame width to frame height);
$n$ = number of frames scanned per second.

The structure of a 441-line picture is sufficiently fine to be unobtrusive when viewed from a comfortable distance. The detail is comparable to that of home movies. A 441-line scanning pattern has therefore been adopted as standard in this country.

The required band width for the transmission of a 441-line image is obtained by substitution in Eq. (1). Thus,

$$\text{frame height } a = 441, \tag{2}$$

$$a^2 = 441^2 = 194{,}481. \tag{3}$$

If the frame width is a third more than the frame height,

$$\text{aspect ratio} = \tfrac{4}{3}. \tag{4}$$

Since, under the present system, 30 frames are transmitted per second,

$$n = 30. \tag{5}$$

Substituting in Eq. (1),

$$\text{required frequency band} = \frac{(194{,}481)\frac{4}{3}(30)}{2}, \tag{6}$$

and

$$\text{frequency band} = 3{,}889{,}620 \tag{7}$$

It has been found that transmission of the foregoing frequency band results in more detail in the direction of the scanning lines than in the vertical direction. Equal detail in both directions has been obtained empirically when the transmission has a flat response to approximately 2.75 megacycles and falls gradually to zero response at 4.25 megacycles. This result is approximated with a system having a band width of approximately 60 per cent that of Eq. (7). For practical purposes, therefore, Eq. (1) is corrected to read

$$\text{required frequency band} = 0.3a^2bn. \tag{8}$$

The frequency band for the 441-line image under discussion would therefore have to be 2,333,772 c. Allowing 10 per cent for synchronizing pulses, the actual band required would be 2,593,080 c. When this is used to modulate a carrier, the total width of the two side bands is thus approximately 5 megacycles.

As a result of the enormous side-band requirements, it has been impossible to utilize the standard broadcast bands for radio television. Experimental television broadcasting has therefore been carried on extremely high frequencies on the order of 40 megacycles or more. On these ultrahigh frequencies, the side-band width is not too great a proportion of the carrier frequency. The FCC recently made a tentative frequency allotment of the u-h-f portion of the spectrum. The television-broadcasting assignments lie between 40 and 85 megacycles, and the frequencies from 150 to 300 megacycles have been tentatively assigned to television relay stations.

It is apparent that the amplifiers used to amplify the iconoscope output currents must be capable of handling very wide frequency range without amplitude or phase distortion. To achieve this condition amplifiers of the impedance-coupled type using resistance-inductance coupling circuits

are usually employed. Special tubes, such as the 1851, 1852, and 1853, have been developed especially for television work and incorporate the interelectrode capacitance and transconductance values necessary for wide-band amplification. In order to distinguish them from r-f and a-f amplifiers, television iconoscope signal amplifiers have been called **video amplifiers.** The output signal obtained from the iconoscope is commonly called a **video-frequency current.**

**Sweep Voltages.** The deflecting voltages, or sweep voltages, applied to the horizontal and vertical deflecting plates of the iconoscope would

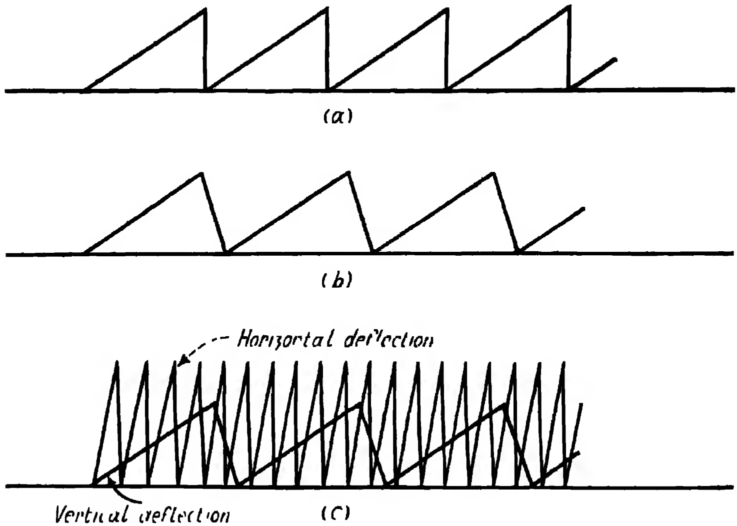



FIG. 314 (a) Theoretical ideal saw-tooth wave form. (b) Typical saw-tooth oscillator-output wave form. (c) Saw-tooth displacement components required for scanning.

be required to have the wave form of a perfect saw tooth—as in Fig. 314(a) for ideal deflection of the electron beam. Such a wave form, of course, is impossible of attainment since it would require that the voltage drop from maximum to zero in a zero time interval. Saw-tooth oscillators have been developed, however, which generate a signal closely approaching the ideal wave form. Such a wave form is shown in Fig. 314(b). The return time, that is, the time required for the voltage to drop from maximum to zero, is reduced to a good minimum.

Saw-tooth voltages of the wave form shown in Fig. 314(b) are generated in television systems by the special form of oscillator illustrated schematically in Fig. 315. The condenser $C$ is charged by voltage $E_b$ through the high resistance $R$. In the absence of an input synchronizing pulse, the tube is nonconducting and does not interfere with the capacitor charging process. With properly chosen circuit parameters, the charging

voltage across $C$ builds up linearly and is coupled to the electron-gun deflecting plates through coupling capacitor $C$. When a synchronizing pulse is applied to the tube input circuit, the negative grid bias is overcome, and plate current flows momentarily. This arrangement affects an almost instantaneous discharge of the capacitor, causing the voltage to drop to zero. When the synchronizing pulse is removed, plate current ceases to flow, and the capacitor starts to charge again, thus beginning another saw-tooth cycle.

One saw-tooth oscillator of this type is coupled to the iconoscope horizontal deflecting plates and another to the vertical deflecting plates. The only difference is in the frequency, the horizontal deflecting frequency being higher than the vertical deflecting frequency. A comparison of the wave forms is shown in Fig. 314(c). During the linear increase of the horizontal deflecting voltage, the electron beam is caused to move from

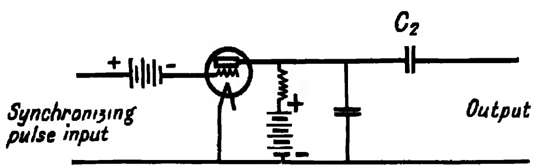



FIG. 315 Saw tooth oscillator used in television systems

left to right across the iconoscope mosaic Upon arriving at the far right-hand side of the mosaic, the voltage drops to zero, causing the electron beam to shift rapidly to the far left-hand side of the mosaic to start another scanning line. At the same time, owing to the more slowly increasing vertical deflection voltage, the beam is caused to move relatively more slowly downward. At the completion of one frame of scanning, the vertical deflecting voltage drops to zero, causing the beam to shift to the top of the mosaic.

The vertical deflecting saw-tooth generator, it is apparent, generates one complete saw-tooth wave form for each image frame and therefore oscillates at a frequency of 30 saw-tooth cycles. The horizontal deflecting saw-tooth generator generates one complete saw tooth wave form for each line of scanning and therefore oscillates at a frequency of $441 \cdot 30$ or 13,230 saw-tooth cycles per second.

**Synchronization.** In television systems synchronization is accomplished by the generation of pulses, one for each scanning line and one for each image frame. The pulses are usually produced by a multivibrator oscillator whose frequency is adjusted by an inductance in the circuit. The pulses are transmitted simultaneously to the proper saw tooth oscillators (Fig. 315) and to the receiver (via modulation of the transmitter carrier). In order to distinguish them from the image signal, the synchronizing pulses are made much greater in amplitude. The

multivibrator providing the line-synchronizing pulses is adjusted to oscillate at scanning-line frequency of 13,230 c. Since it also triggers the scanning-line saw tooth oscillator, this pulse occurs during the period in which the scanning beam returns to starting (left-hand) position; and since it is greater in amplitude than the image signal caused by a black portion of the image, this pulse also effectively blacks out the return trace of the beam in the receiver viewing tube (kinescope).

The multivibrator providing the frame-synchronizing pulse is adjusted to oscillate at frame frequency, or 30 c per second. Since this pulse simultaneously triggers the vertical deflecting saw-tooth oscillator, it is considerably longer in duration and is therefore easily distinguishable from the line pulse at the receiver. This frame pulse also functions to obliterate the upward return trace of the beam in the receiver viewing tube.

**Flicker.** It has been found that, although the eye is unaware of discontinuity of motion at frequencies above 16 c per second, it nevertheless can detect flicker or change in light intensity at very much higher frequencies. The threshold frequency for awareness of flicker is approximately 48 c under average conditions. The exact value depends upon a great many factors, including brightness of the object, color of the light, relative duration of light and dark, and so on. The flicker threshold of 48 c has been recognized by the motion-picture industry, and it is the reason for the standard frame frequency of 24 c. Flicker is avoided by momentarily blocking out the exciting lamp in the middle of each frame. The light consequently flashes on the screen twice for each frame. As a result, the image is presented to the observer 48 times per second, and the eye is unaware of flicker.

Because of the fact that stray power-frequency currents unavoidably creep into television amplifier and deflection oscillator circuits, if a 24-c frame frequency were used in television, periodic flicker and ripples in the received image would be created by the 60-c strays. To eliminate this disturbance, a frame frequency of 30 c was chosen as standard. Since 60-c power frequencies are almost universal in this country, any induced 60-c stray voltage will coincide with the frame frequency, and no objectionable image interference will result.

Unfortunately, however, a frame frequency of 30 c is still below the eye flicker awareness threshold of 48 c per second. This difficulty is overcome by transmitting half an image at a time. The effective image presentation at the receiver consequently occurs at a frequency of 60 c per second, and no flicker is detectable. Half-image transmission is accomplished by *interlaced scanning*, illustrated in Fig. 316. The foregoing discussion has been based on a system utilizing *sequential scanning*, that is, scanning in which the lines are scanned progressively from top to bottom in the mosaic. In interlaced scanning, every other line is skipped

as the beam progresses down the mosaic. After the frame has once been scanned in this fashion, the intervening spaces (skipped lines) are scanned. In Fig. 316 the figures to the right of the drawing illustrate the order in which the lines are scanned.

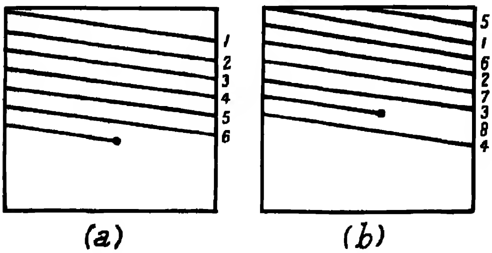



FIG. 316. (a) Sequential scanning (b) Interlaced scanning.

Interlaced scanning is accomplished by adjustment of the relative sawtooth oscillator frequencies and by prolonging the sweep return time interval of the horizontal oscillator

**The Television Receiver.** All modern television receiving systems utilize a cathode-ray tube as the major image-receiving element. The tubes used for this application are similar to the conventional cathode ray tubes employed in oscilloscope work Television image receiving tubes, however, are called **kinescopes** in some systems and, in general, are much larger than conventional cathode ray tubes in order to permit

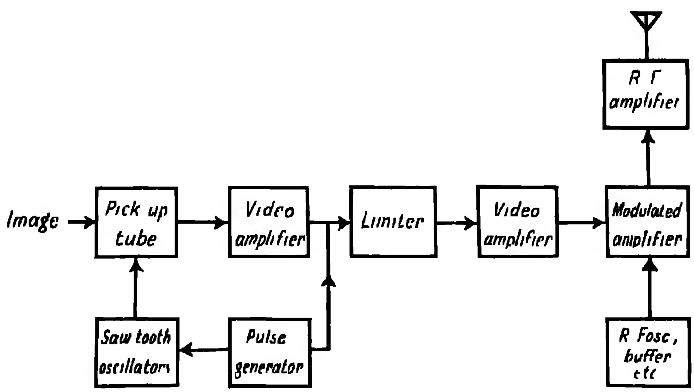



FIG. 317. Block diagram of television transmitter system

fair-sized images to be reproduced. Higher anode voltages are also employed to produce higher levels of illumination.

In the kinescope, the beam from an electron gun is deflected by vertical and horizontal deflecting plates to produce symmetrical scanning of a fluorescent coating on the wall of the tube envelope. Most modern tubes utilize willemite for the fluorescent material. When the electron beam impinges upon the willemite screen, it fluoresces, the intensity of the light developed depending upon the magnitude of the electron stream.

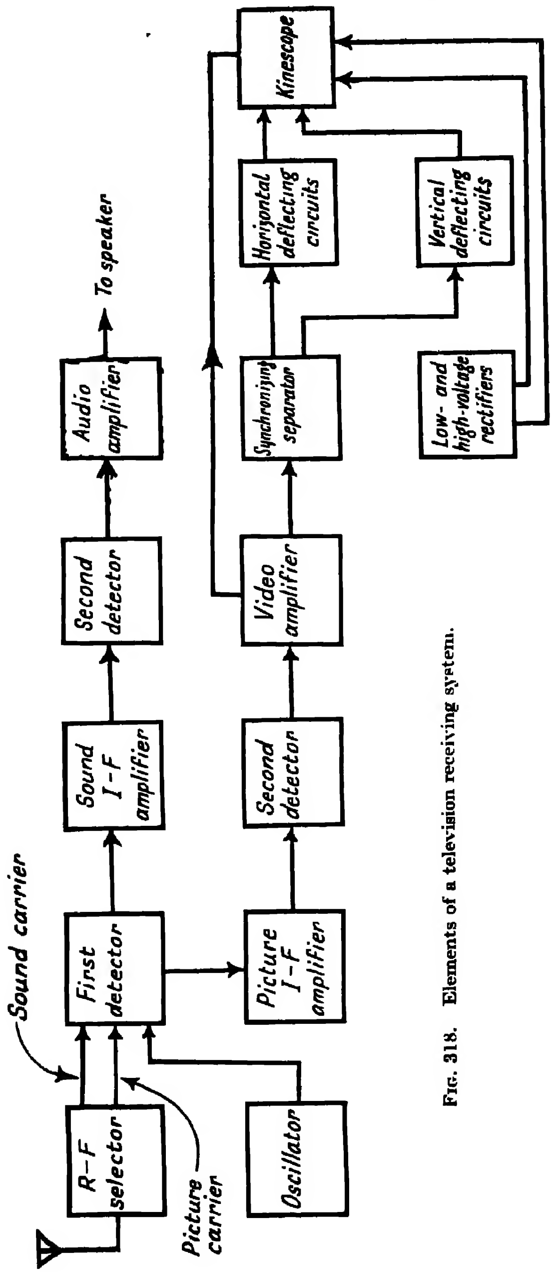



FIG. 318. Elements of a television receiving system.

The television receiver utilizes a superheterodyne circuit especially designed to pass the wide frequency bands that are necessary. Usually an audio receiver (another superheterodyne) is coupled to a common input circuit, thus affording simultaneous reception of both the aural and the visual signal. A common h f oscillator is employed to produce the

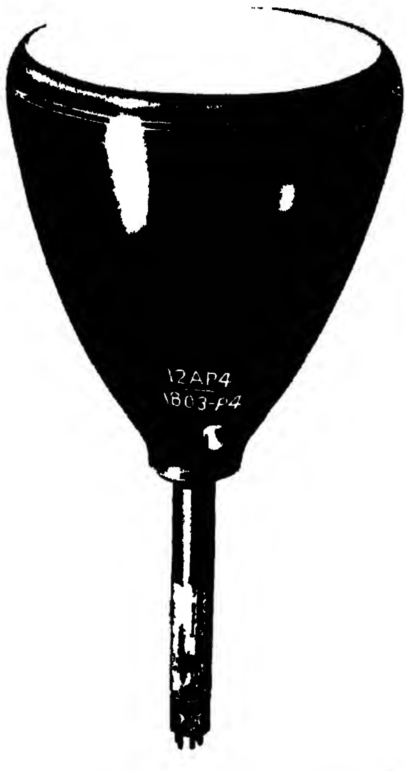



FIG 319 Typical kinescope television tube (Courtesy of RCA Manufacturing Co, Inc)

i-f heterodyne for both receivers. The video receiver intermediate frequency is usually on the order of 5 to 10 megacycles and utilizes band pass coupling circuits to provide the necessary linear response over the desired wide range. The sound program is generally transmitted by a separate carrier The sound receiver utilizes a comparatively narrow i-f band-pass circuit Because of the common h-f oscillator, the sound receiver can be used as a pilot to determine when the video receiver is correctly tuned to its carrier

After the second detector of the video receiver, the signal is amplified

by a video amplifier that brings the video signal level up to the value necessary to operate the kinescope. This signal is then impressed on the kinescope grid and causes the electron beam to vary in intensity.

Two local saw-tooth oscillators generate the horizontal and vertical deflecting plate voltages in a manner similar to that used at the transmitter. Proper timing is assured by the synchronizing pulses that are received with the video signal in the second-detector plate circuit and utilized to trigger the oscillator tubes. The synchronizing pulses are separated from the video signal by coupling them to an amplifier that is biased considerably beyond cutoff. Only signals above the cutoff level (which is preadjusted above the video signal level) cause plate current to flow in the amplifier. The line and frame synchronizing pulses are separated by virtue of the difference in frequency and duration. This is accomplished by means of simple high- and low-pass filter circuits consisting of resistance and capacitance. The horizontal pulse is isolated by a series capacitance and shunt resistance in the input of the horizontal saw-tooth oscillator. The vertical pulse is isolated by a series resistance and shunt capacitance in the vertical oscillator input circuit.

## *RADIO FACSIMILE*

The potentialities of facsimile as a means of expanding the capabilities of existing radio communication systems have served to focus attention on facsimile application in the field of radio. Facsimile, in general, is a means of converting written or printed copy, photographs, and other types of illustration into electric signals that may be transmitted via conventional wire or radio circuits. The technical features of facsimile transmission are similar, in general, to those of the widely used wirephoto and telephotograph circuits.

**General Theory.** A complete radio facsimile system comprises a scanning mechanism coupled to a conventional radiotelephone transmitter and a recording mechanism coupled to a conventional receiver output circuit. Copy to be transmitted is placed in a curved *scanning gate*. The copy is fed upward from the scanning gate to a feed roller that is operated by a pawl and ratchet arrangement. Light from a powerful exciter lamp is focused through a lens barrel upon a spot 0.01 in. in diameter on the copy. By synchronously controlled movement of the lens system, the light spot traces, or scans, a line from left to right across the copy. After each left-to-right trace, the copy is moved upward 0.01 in., during which time the scanning lens returns from right to left, or starting, position. During this return movement of the scanning head, no picture signal is transmitted.

During the left-to-right scanning trace, light from the illuminated 0.01-in. spot on the copy is reflected into a photoelectric cell of the cesium

cathode type through another lens arrangement mounted at a downward angle from the scanning head. The photocell output is used to modulate a 2,000-c oscillator. The output of this oscillator is fed into the conventional speech-amplifier circuit of the radio transmitter proper.

Maximum light is reflected from white copy, minimum from black. The photoelectric cell is so coupled to the audio-modulator unit that black copy results in maximum amplitude of the 2,000-c oscillator and white copy produces minimum amplitude. Intervening gradations of color between black and white produce corresponding variations in the audio-oscillator amplitude.

At the receiving end of the system, the a-f output of the receiver consists of an intermittent 2,000-c note varying in amplitude. This audio output is rectified, and the resulting direct current is applied to a recording stylus. The latter consists of fine tungsten wire 0.01 in. in diameter mounted in a tubular sleeve and so disposed as to bear against the paper upon which the recorded copy is traced. The lateral movement of the stylus and upward feed movement of the copy paper are synchronized with the scanning-unit mechanism at the transmitter. The method of synchronism is discussed in a following section.

The recording paper is of a special electrosensitive material. The surface of the paper instantly darkens when a marking voltage of sufficient intensity is applied to the stylus. Marking of the paper usually commences at approximately 80 v (stylus to ground) and increases in density with increasing voltage. At approximately 170 v, a jet black trace of maximum density is attained. Since the d-c rectifier output voltage (the voltage applied to the stylus) varies directly with the amplitude of the 2,000-c audio-signal component, the result is a reproduction of the transmitted copy on the electrosensitized paper.

**The Scanner.** All the mechanical operations in the scanning unit are performed by an 1,800-rpm synchronous motor. Through reduction gears and the previously mentioned pawl and ratchet arrangement the upward movement of the copy is derived. The copy, which may be in rolls several hundred feet long, is pressed firmly against the knurled surface of a feed roller actuated by the ratchet mechanism. During each return stroke of the scanning head, the ratchet gear effects a 0.01-in. upward movement of the feed roller and, hence, the copy through a cam movement coupling it to the driving motor.

The scanning head comprises an integral assembly of an exciter lamp, photocell, and two lens systems. One lens system focuses the light from the exciting lamp through a small lateral aperture called the "scanning gate" upon a spot on the copy 0.01 in. in diameter. The other lens system focuses the light reflected from this tiny spot upon the photocell.

The entire scanning-head assembly is given a reciprocating, or to-and-fro, motion by means of another cam arrangement. The reciprocating

motion results in a lateral movement of the light spot across the copy, from left to right, and return. Only the reflected light from the left-to-right movement is utilized for transmission purposes. During the right-to-left, or return, movement, the copy is moved upward 0.01 in., and the scanning head is returned to the starting position to start a new scanning cycle over a new portion of the copy 0.01 in. below the previously scanned portion.

It is apparent that a variation in synchronism between transmitting and receiving equipment could occur if the respective driving motors did not operate at constant speeds. Such a nonsynchronous system could result in serious distortion of the received copy if the variation, either lead or lag, were permitted to become cumulative. Such distortion is minimized by using motors that are, so far as possible, constant-speed units. Where an alternating current of stable frequency is available, synchronous motors are employed. In installations where it is necessary to utilize d-c motors, such as aboard ship or in mobile installations ashore, speed variation is minimized by the use of governors. In addition to the above precautions, effective synchronism is further ensured by the transmission of a *synchronizing pulse*. This pulse, consisting of a 500-c signal, is transmitted during the time interval in which the scanning head returns to starting position and ensures simultaneous starting of both the scanning head and recording stylus at the beginning of each cycle. Any error introduced by relative speed variation between scanner and recorder motors is therefore corrected at the beginning of each scanning cycle, and resultant distortion is prevented from becoming cumulative and is kept at a minimum.

The circuit of a typical scanner is shown in Fig. 320. The 2,000-c facsimile carrier is provided by a conventional push-pull a-f oscillator utilizing a 6E6 tube. During the scanning stroke, this oscillator is coupled to the screen grid of a type 1851 tube, which acts as an a-f modulator. Light variations on the photocell cathode produce changes in control-grid voltage of the 1851. This results in a variation in amplitude of the carrier signal present in the plate circuit of the modulator tube.

During the scanning operation, the plate circuit of the modulator is coupled to the input circuit of a triode (6C5) signal amplifier through a switch operated by a cam on one of the motor-driven shafts. The triode amplifier output is fed into the conventional speech-amplifier circuits of a radiotelephone transmitter.

During the return, or inoperative, period of the scanning cycle, a number of switching operations occur, all of which are controlled by motor-driven cams arranged to provide the proper timing sequence. As soon as the scanning (left to right) movement is completed, a switch disconnects the triode amplifier input from the modulator output circuit and connects it to one side of a synchronizing pulse switch. At the same

time, a second switch shunts additional capacity across the 6E6 oscillator tank circuit, shifting the oscillator frequency from 2,000 to 500 c. Just before the scanning head commences the new operating cycle, the synchronizing-pulse switch closes momentarily, connecting the oscillator, now operating at 500 c, to the triode amplifier and thence to the radio-transmitter circuits. The 500-c synchronizing pulse is thus transmitted

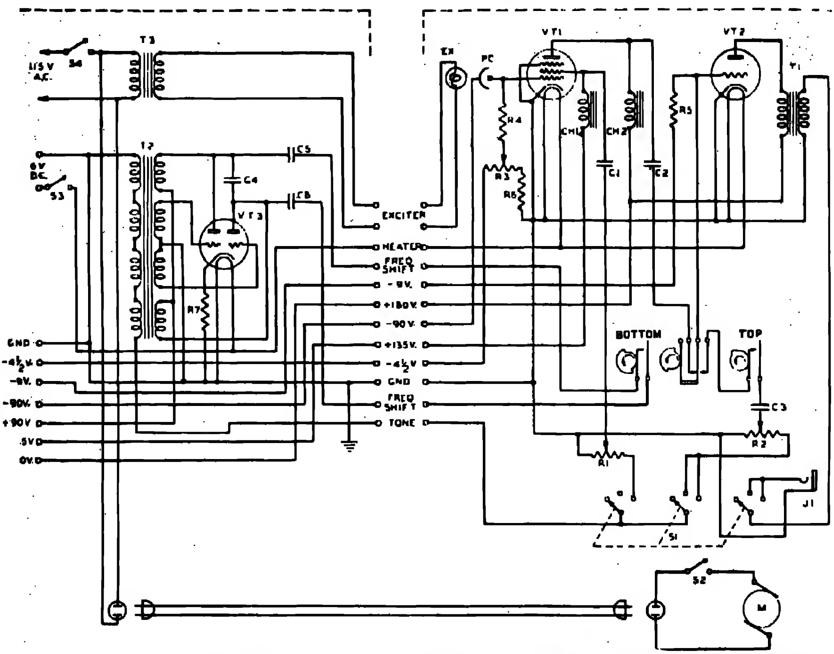



FIG. 320. Circuit diagram of two-column scanner. (*Courtesy of Finch Tele-communications Laboratories, Inc.*)

for a period of approximately one fiftieth of a second. Immediately thereafter, all switches return to normal position by cam action: the synchronizing-pulse switch is open; the oscillator is switched back to 2,000-c operation; and the triode amplifier is connected back to the modulator output circuit. The scanning operation then proceeds as in the previous cycle.

**The Recorder.** In most of its mechanical aspects, the facsimile recorder is similar to the scanner. The upward movement of the copy paper is controlled by a pawl and ratchet mechanism similar to that at the scanner. The reciprocating movement of the recording arm is derived from cam action.

The recording stylus, as previously mentioned, is a tungsten wire 0.01 in. in diameter and is supported by a small tubular sleeve. The

stylus and sleeve, in turn, are flexibly mounted in a cradle at the end of the recording arm. The tip of the stylus is held against the paper by a delicate spring, the tension of which is arranged to permit self-adjustment

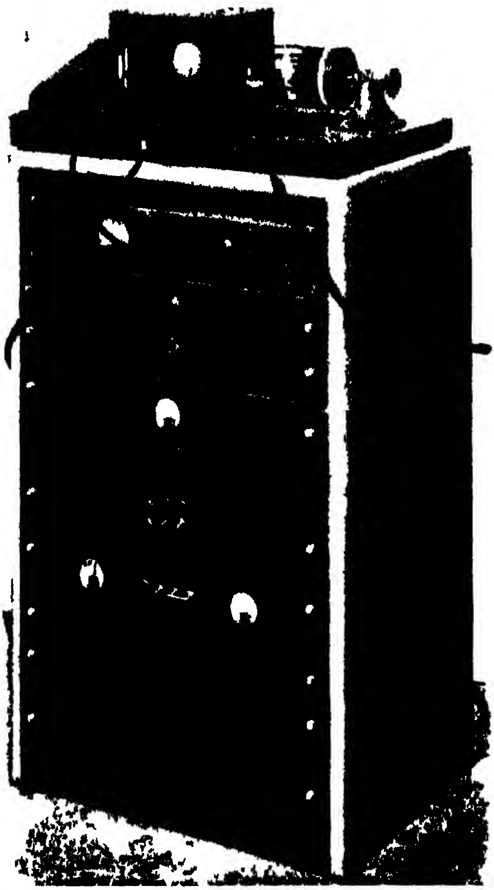


FIG. 321(a) Commercial model facsimile scanner (*Courtesy of Finch Telecommunications, Inc.*)

of the stylus for irregularities and imperfections in the coated surface of the paper.

Proper sequence and timing of the various mechanical operations are obtained by means of a simple two-disk friction clutch which couples the driving cams to the motor gear reduction box. Once each revolution, the cam rotations are completely stopped when a projecting segment of the clutch-driven disk encounters a stop lever. The stop lever is attached

to the armature of an electromagnet that is actuated by the 500-c autosynchronizing pulse.

At the stop position, the recording arm is in position at the starting position, that is, at the extreme left-hand side of the recording gate. When the 500-c synchronizing pulse actuates the electromagnet, the stop lever is drawn back, thus releasing the clutch disk and permitting it to

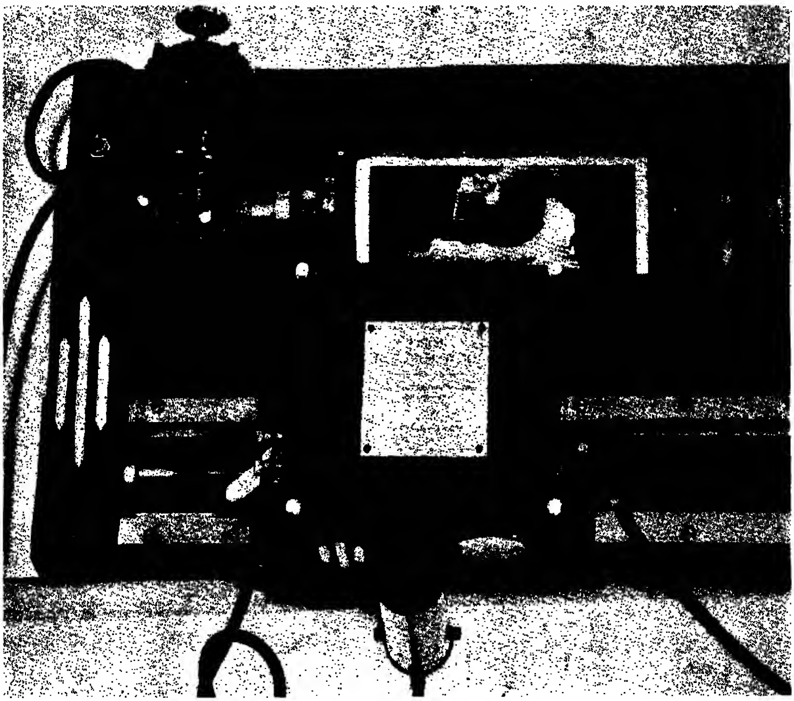


FIG. 321(b). Close-up of scanning head and exciter lamp housing. (*Courtesy of Finch Telecommunications, Inc.*)

engage the drive disk. All cam operations thereupon immediately commence and the recording arm moves across the paper from left to right in synchronism with the scanning head at the transmitter. Simultaneously, the recording paper is fed through the scanning gate in an upward direction. There is no movement of the paper unless the recording arm is in motion.

In order to permit correction of any discrepancy in timing caused by motor speed variation, the recording cycle is completed in a fraction of a second less time than that required for the scanning cycle. This is accomplished by decreasing the effective speed of the cam operations in the recorder. In units employing synchronous motors, this reduction is

accomplished by different gearing. In d-c installations the motor speed itself may be preadjusted. In this manner, the entire recording operating cycle from left to right and return is completed just before the scanning cycle at the transmitter has been completed. At this point, owing to the action of the clutch stop, all recording operations momentarily cease. Just before the completion of the scanning cycle at the transmitter, the

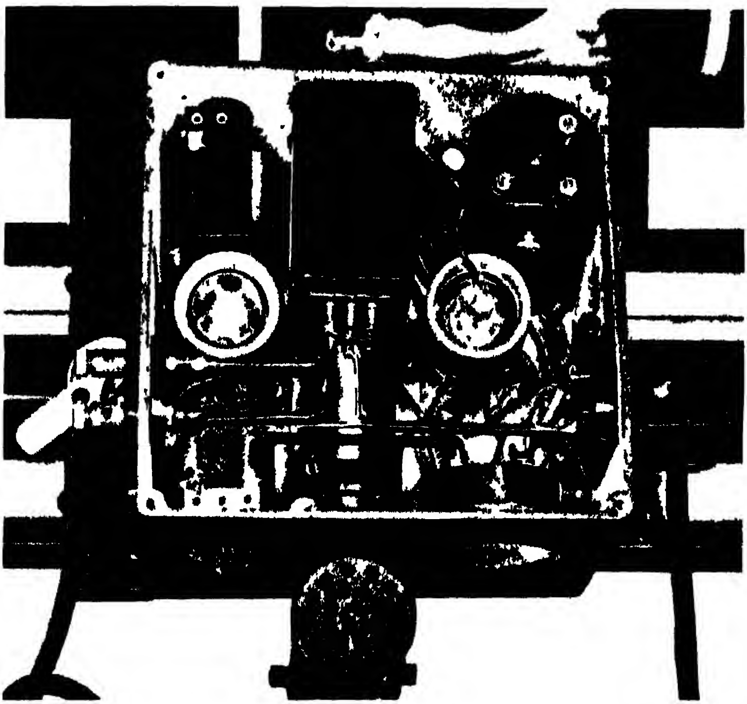


FIG. 321(c). Interior of scanning head. (*Courtesy of Finch Telecommunications, Inc.*)

500-c synchronizing pulse is transmitted, releasing the clutch through electromagnetic action. The beginning of the recording cycle then coincides precisely with the beginning of the scanning cycle. During the stop interval in the recorder cycle any discrepancy in synchronism may be corrected. Such discrepancies therefore appear in the recorder simply as a variation in the stop interval instead of as a variation in stylus progress across the copy with resultant distortion.

The electric circuit of a radio facsimile recorder is shown in Fig. 322. It will be seen that the electronic portion of the circuit consists of a conventional full-wave rectifier utilizing a 6A6 or 25Z6 tube. The rectifier transformer is coupled directly to the a-f output circuit of a conventional

receiver. A step-up transformer is customarily utilized to raise the a-f a-c signal voltage to the potential necessary for effective marking.

During the recording, or left to-right, movement of the recording arm, a d-c marking voltage is applied to the stylus from the 6A6 rectifier through the cam-switch contacts. Shortly before the recording arm reaches the start position on the return stroke of the cycle, the cam switch transfers the rectifier output to the synchronizing electromagnet. Simultaneously, another section of this cam switch connects a by-pass

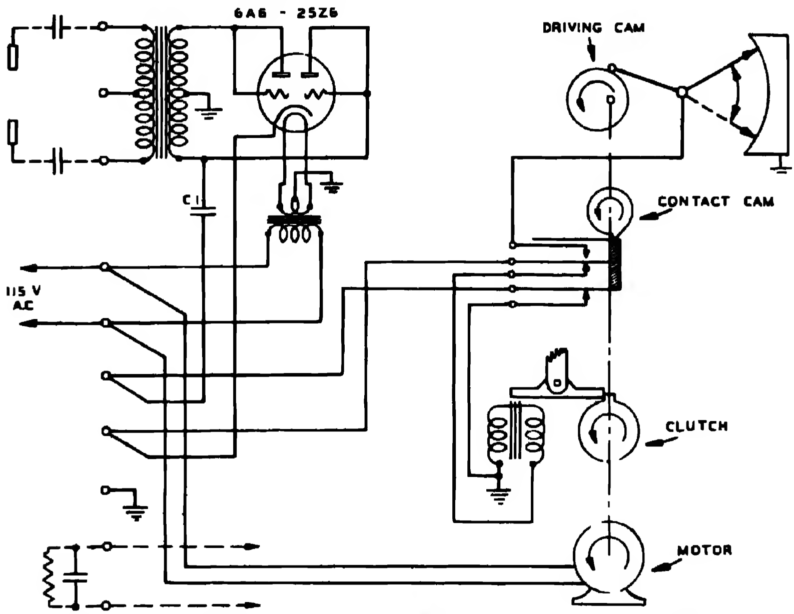



FIG. 322. Circuit diagram of two column recorder (Courtesy of Finch Telecommunications, Inc.)

condenser $C_1$ across half of the transformer secondary. In this position, the cam operations are momentarily halted by the action of the clutch stop lever. When the 500-c synchronizing pulse is received and rectified, it passes through the electromagnet winding, attracting the armature and releasing the clutch. The subsequent cam rotation connects all switch circuits back to normal, or recording, position.

The condenser $C_1$ is connected across half the transformer secondary during the stop interval to discriminate against signals higher in frequency than the 500-c synchronizing pulse. In addition to discriminating against extraneous noise voltages, which might conceivably cause false operation of the synchronizing electromagnet, this capacitor by-passes the 2,000-c facsimile carrier, preventing it from releasing the stop lever. For proper operation, a receiver having an a-f power output of at least 3 w is required.

The factors of brightness, contrast, and so on, in the received facsimile copy are dependent to a great extent upon the type and quality of the recording paper utilized. In the system under discussion, a dry carbon-impregnated stock is used on which a near-white or orange-colored coating of electrosensitive material has been applied. The paper is customarily supplied in rolls of 100 ft or more for the small-width type of machine. Because the width is equivalent to the width of two standard newspaper columns, the unit is usually called a **two-column recorder.**

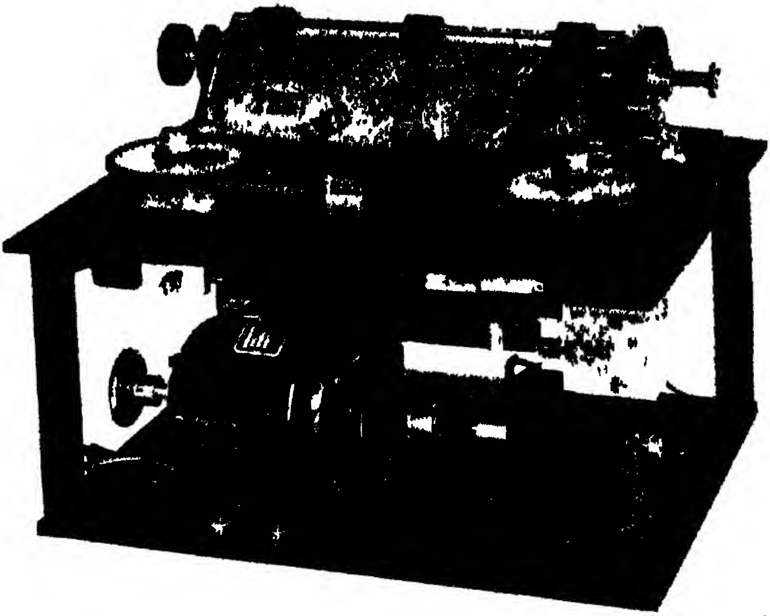


FIG. 323 Commercial model facsimile recorder utilizing three styluses (*Courtesy of Finch Telecommunications, Inc.*)

Facsimile equipment is at present available in widths from two to five columns. As the art progresses, the limitations of width, transmitting speed, and so on, will undoubtedly be greatly extended. With a five-column unit, it is possible to print five-column copy at the rate of 22 sq in. per minute, or 10 ft per hour. This is the equivalent of eight tabloid newspaper pages per hour. The possibilities of radio facsimile as a medium for the transmission of printed or written intelligence are at once apparent. Thus, a facsimile unit that scans at the rate of 8 sq in. per minute, 100 lines per inch, is capable of transmitting single-spaced type-written copy of standard telegraph blank size at a speed equivalent to 150 words per minute. By utilizing smaller-size type and greater-width

facsimile machines, much higher speeds are possible. As the type size is decreased, however, the possibility of mutilations caused by fading, by extraneous noise voltage due to atmospherics, and so on, becomes greater. Nevertheless, the permissible margin of error of this kind is much greater than that experienced in conventional telegraph transmission, either hand sending or high-speed automatic sending. Thus, a crash of static could easily totally obliterate several letters or perhaps an entire word of high-speed automatic-telegraph transmission. A static crash of similar duration would be evidenced in a facsimile recording as a lateral streak in the copy. The copy would still be legible

A number of schemes have been proposed for the adaptation of facsimile transmission to radiotelegraph circuits. One plan would permit the

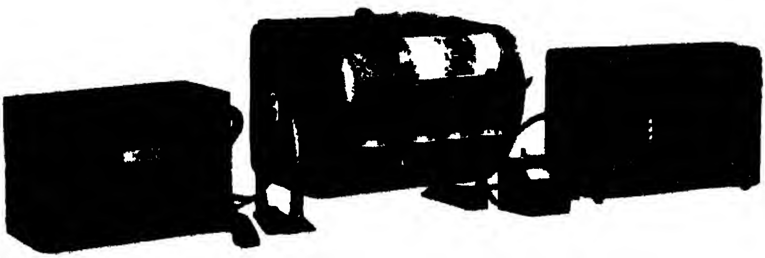


FIG. 324. Complete facsimile duplex unit showing power supply at left and amplifier at extreme right. (*Courtesy of Finch Telecommunications, Inc.*)

transmission of a given area of copy a standard telegram blank, for example—for a flat rate, regardless of how much was written upon it. The telegram would then be delivered in the sender's handwriting and could even include a sketch. The system could also be used for the transmission of technical drawings, such as blueprints. Because of legal and economic difficulties, however, no commercial radio facsimile service of this kind has as yet been developed

An interesting application of radio facsimile is its use by radio broad casting stations for the dissemination of news bulletins and current news photographs. A number of stations throughout the country now provide this service, which consists of the transmission of a radio newspaper that is literally printed in the subscriber's home. Relatively inexpensive facsimile recorders that may readily be connected to the average medium-power home radio are available to the public. After the close of the normal broadcast day, a complete news bulletin is transmitted by facsimile from the local broadcast station, usually from midnight to 6 A.M. Before retiring, the subscriber sets his radio with the recorder switched

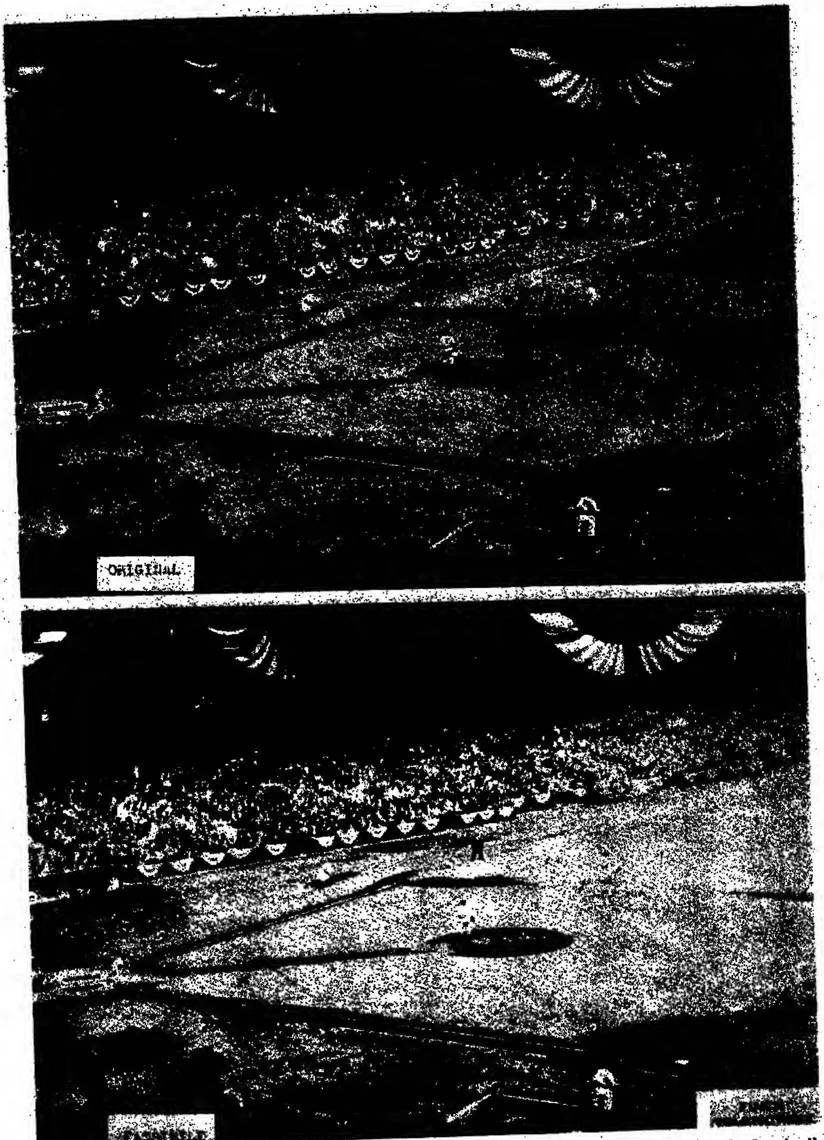



FIG. 325. Example of a photograph radio-transmitted by means of facsimile. (*Courtesy of Finch Telecommunications, Inc.*)

in. The next morning, a complete printed and illustrated news bulletin awaits him. Electric timing switches are also available which turn the equipment on and off at specified times.

For portable communications applications, duplex radio facsimile units have been developed. Such units incorporate the functions of scanner and recorder in a single unit and can be employed with a conventional radiotelephone transmitter and receiver. A typical duplex facsimile unit is shown in Fig. 324.

## *QUESTIONS AND PROBLEMS**

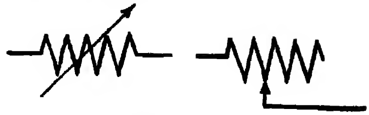
1. What is the basic difference between television and facsimile?
2. Define aspect ratio.
3. Explain the principle of operation of the iconoscope.
4. It is desired to transmit a television image of 400 lines with an aspect ratio of 4 to 3. Thirty frames per second will be scanned. What is the required frequency band width?
5. Explain how scanning line pulses and image-frame pulses are distinguished from each other in television synchronization systems.
6. Why has a frame frequency of 30 c been standardized for television stations in the United States?
7. What is the advantage of interlaced scanning?
8. How is light intensity converted into electrical energy in a facsimile scanning system?
9. Draw a simple diagram of a facsimile scanner circuit.
10. How is synchronism accomplished between the scanner and the recorder in a facsimile system?

* These questions and problems are taken from the "F C C Study Guide for Commercial Radio Operator Examinations."

# APPENDIX

TABLE I. Schematic Radio Symbols

Fixed Resistor

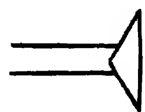


Variable Resistor

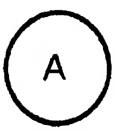
Wires Connected

Wires Crossed but Not Connected

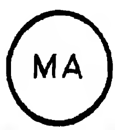
Headphones



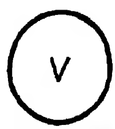
Loudspeaker



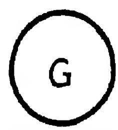

Ammeter




Milliammeter

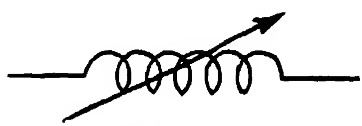



Voltmeter

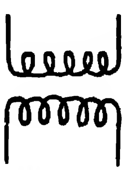



Galvanometer

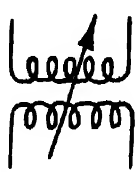
Fixed Air-core Inductor
(R-F Choke)



Variable Air-core Inductor



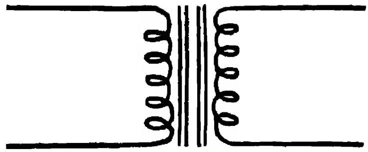
Air-core Transformer, Fixed Coupling
(R-F or I-F Transformer)



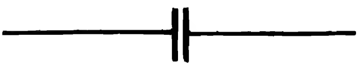
Air-core Transformer, Variable
Coupling

TABLE I (contd )

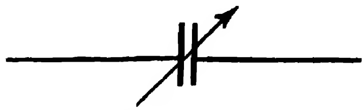
Iron core Inductor (A F or Filter Choke)



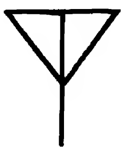
Iron core Transformer (A F or Power Transformer)



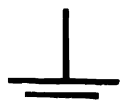
Fixed Capacitor



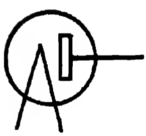
Variable Capacitor



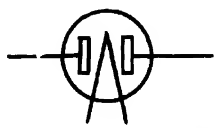
Antenna



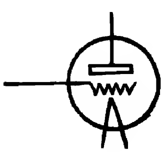
Ground



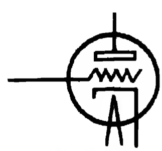
Diode (Half wave Rectifier)



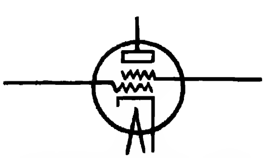
Duo diode (Full wave Rectifier)



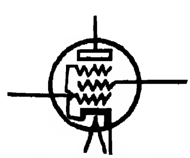
Triode



Heater type Triode

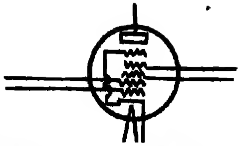


Tetrode (Screen grid Tube)

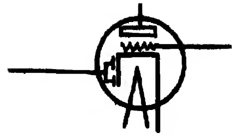


Pentode (Suppressor grid Tube)

**Table I** (*contd.*)



Pentagrid Converter (Used in Superheterodyne Mixer Circuits)



Duo-diode High-mu Triode (Combination Diode Detector and First Audio-amplifier Tube)

**Table II. Mantissas of Common Logarithms**

| *N* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 0000 | 0043 | 0086 | 0128 | 0170 | 0212 | 0253 | 0294 | 0334 | 0374 |
| 11 | 0414 | 0453 | 0492 | 0531 | 0569 | 0607 | 0645 | 0682 | 0719 | 0755 |
| 12 | 0792 | 0828 | 0864 | 0899 | 0934 | 0969 | 1004 | 1038 | 1072 | 1106 |
| 13 | 1139 | 1173 | 1206 | 1239 | 1271 | 1303 | 1335 | 1367 | 1399 | 1430 |
| 14 | 1461 | 1492 | 1523 | 1553 | 1584 | 1614 | 1644 | 1673 | 1703 | 1732 |
| 15 | 1761 | 1790 | 1818 | 1847 | 1875 | 1903 | 1931 | 1959 | 1987 | 2014 |
| 16 | 2041 | 2068 | 2095 | 2122 | 2148 | 2175 | 2201 | 2227 | 2253 | 2279 |
| 17 | 2304 | 2330 | 2355 | 2380 | 2405 | 2430 | 2455 | 2480 | 2504 | 2529 |
| 18 | 2553 | 2577 | 2601 | 2625 | 2648 | 2672 | 2695 | 2718 | 2742 | 2765 |
| 19 | 2788 | 2810 | 2833 | 2856 | 2878 | 2900 | 2923 | 2945 | 2967 | 2989 |
| 20 | 3010 | 3032 | 3054 | 3075 | 3096 | 3118 | 3139 | 3160 | 3181 | 3201 |
| 21 | 3222 | 3243 | 3263 | 3284 | 3304 | 3324 | 3345 | 3365 | 3385 | 3404 |
| 22 | 3424 | 3444 | 3464 | 3483 | 3502 | 3522 | 3541 | 3560 | 3579 | 3598 |
| 23 | 3617 | 3636 | 3655 | 3674 | 3692 | 3711 | 3729 | 3747 | 3766 | 3784 |
| 24 | 3802 | 3820 | 3838 | 3856 | 3874 | 3892 | 3909 | 3927 | 3945 | 3962 |
| 25 | 3979 | 3997 | 4014 | 4031 | 4048 | 4065 | 4082 | 4099 | 4116 | 4133 |
| 26 | 4150 | 4166 | 4183 | 4200 | 4216 | 4232 | 4249 | 4265 | 4281 | 4298 |
| 27 | 4314 | 4330 | 4346 | 4362 | 4378 | 4393 | 4409 | 4425 | 4440 | 4456 |
| 28 | 4472 | 4487 | 4502 | 4518 | 4533 | 4548 | 4564 | 4579 | 4594 | 4609 |
| 29 | 4624 | 4639 | 4654 | 4669 | 4683 | 4698 | 4713 | 4728 | 4742 | 4757 |
| 30 | 4771 | 4786 | 4800 | 4814 | 4829 | 4843 | 4857 | 4871 | 4886 | 4900 |
| 31 | 4914 | 4928 | 4942 | 4955 | 4969 | 4983 | 4997 | 5011 | 5024 | 5038 |
| 32 | 5051 | 5065 | 5079 | 5092 | 5105 | 5119 | 5132 | 5145 | 5159 | 5172 |
| 33 | 5185 | 5198 | 5211 | 5224 | 5237 | 5250 | 5263 | 5276 | 5289 | 5302 |
| 34 | 5315 | 5328 | 5340 | 5353 | 5366 | 5378 | 5391 | 5403 | 5416 | 5428 |
| 35 | 5441 | 5453 | 5465 | 5478 | 5490 | 5502 | 5514 | 5527 | 5539 | 5551 |
| 36 | 5563 | 5575 | 5587 | 5599 | 5611 | 5623 | 5635 | 5647 | 5658 | 5670 |
| 37 | 5682 | 5694 | 5705 | 5717 | 5729 | 5740 | 5752 | 5763 | 5775 | 5786 |
| 38 | 5798 | 5809 | 5821 | 5832 | 5843 | 5855 | 5866 | 5877 | 5888 | 5899 |
| 39 | 5911 | 5922 | 5933 | 5944 | 5955 | 5966 | 5977 | 5988 | 5999 | 6010 |
| 40 | 6021 | 6031 | 6042 | 6053 | 6064 | 6075 | 6085 | 6096 | 6107 | 6117 |
| 41 | 6128 | 6138 | 6149 | 6160 | 6170 | 6180 | 6191 | 6201 | 6212 | 6222 |
| 42 | 6232 | 6243 | 6253 | 6263 | 6274 | 6284 | 6294 | 6304 | 6314 | 6325 |
| 43 | 6335 | 6345 | 6355 | 6365 | 6375 | 6385 | 6395 | 6405 | 6415 | 6425 |
| 44 | 6435 | 6444 | 6454 | 6464 | 6474 | 6484 | 6493 | 6503 | 6513 | 6522 |
| 45 | 6532 | 6542 | 6551 | 6561 | 6571 | 6580 | 6590 | 6599 | 6609 | 6618 |
| 46 | 6628 | 6637 | 6646 | 6656 | 6665 | 6675 | 6684 | 6693 | 6702 | 6712 |
| 47 | 6721 | 6730 | 6739 | 6749 | 6758 | 6767 | 6776 | 6785 | 6794 | 6803 |
| 48 | 6812 | 6821 | 6830 | 6839 | 6848 | 6857 | 6866 | 6875 | 6884 | 6893 |
| 49 | 6902 | 6911 | 6920 | 6928 | 6937 | 6946 | 6955 | 6964 | 6972 | 6981 |
| 50 | 6990 | 6998 | 7007 | 7016 | 7024 | 7033 | 7042 | 7050 | 7059 | 7067 |
| 51 | 7076 | 7084 | 7093 | 7101 | 7110 | 7118 | 7126 | 7135 | 7143 | 7152 |
| 52 | 7160 | 7168 | 7177 | 7185 | 7193 | 7202 | 7210 | 7218 | 7226 | 7235 |
| 53 | 7243 | 7251 | 7259 | 7267 | 7275 | 7284 | 7292 | 7300 | 7308 | 7316 |
| 54 | 7324 | 7332 | 7340 | 7348 | 7356 | 7364 | 7372 | 7380 | 7388 | 7396 |

TABLE II (*contd*)

| *N* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 55 | 7404 | 7412 | 7419 | 7427 | 7435 | 7443 | 7451 | 7459 | 7466 | 7474 |
| 56 | 7482 | 7490 | 7497 | 7505 | 7513 | 7520 | 7528 | 7536 | 7543 | 7551 |
| 57 | 7559 | 7566 | 7574 | 7582 | 7589 | 7597 | 7604 | 7612 | 7619 | 7627 |
| 58 | 7634 | 7642 | 7649 | 7657 | 7664 | 7672 | 7679 | 7686 | 7694 | 7701 |
| 59 | 7709 | 7716 | 7723 | 7731 | 7738 | 7745 | 7752 | 7760 | 7767 | 7774 |
| 60 | 7782 | 7789 | 7796 | 7803 | 7810 | 7818 | 7825 | 7832 | 7839 | 7846 |
| 61 | 7853 | 7860 | 7868 | 7875 | 7882 | 7889 | 7896 | 7903 | 7910 | 7917 |
| 62 | 7924 | 7931 | 7938 | 7945 | 7952 | 7959 | 7966 | 7973 | 7980 | 7987 |
| 63 | 7993 | 8000 | 8007 | 8014 | 8021 | 8028 | 8035 | 8041 | 8048 | 8055 |
| 64 | 8062 | 8069 | 8075 | 8082 | 8089 | 8096 | 8102 | 8109 | 8116 | 8122 |
| 65 | 8129 | 8136 | 8142 | 8149 | 8156 | 8162 | 8169 | 8176 | 8182 | 8189 |
| 66 | 8195 | 8202 | 8209 | 8215 | 8222 | 8228 | 8235 | 8241 | 8248 | 8254 |
| 67 | 8261 | 8267 | 8274 | 8280 | 8287 | 8293 | 8299 | 8306 | 8312 | 8319 |
| 68 | 8325 | 8331 | 8338 | 8344 | 8351 | 8357 | 8363 | 8370 | 8376 | 8382 |
| 69 | 8388 | 8395 | 8401 | 8407 | 8414 | 8420 | 8426 | 8432 | 8439 | 8445 |
| 70 | 8451 | 8457 | 8463 | 8470 | 8476 | 8482 | 8488 | 8494 | 8500 | 8506 |
| 71 | 8513 | 8519 | 8525 | 8531 | 8537 | 8543 | 8549 | 8555 | 8561 | 8567 |
| 72 | 8573 | 8579 | 8585 | 8591 | 8597 | 8603 | 8609 | 8615 | 8621 | 8627 |
| 73 | 8633 | 8639 | 8645 | 8651 | 8657 | 8663 | 8669 | 8675 | 8681 | 8686 |
| 74 | 8692 | 8698 | 8704 | 8710 | 8716 | 8722 | 8727 | 8733 | 8739 | 8745 |
| 75 | 8751 | 8756 | 8762 | 8768 | 8774 | 8779 | 8785 | 8791 | 8797 | 8802 |
| 76 | 8808 | 8814 | 8820 | 8825 | 8831 | 8837 | 8842 | 8848 | 8854 | 8859 |
| 77 | 8865 | 8871 | 8876 | 8882 | 8887 | 8893 | 8899 | 8904 | 8910 | 8915 |
| 78 | 8921 | 8927 | 8932 | 8938 | 8943 | 8949 | 8954 | 8960 | 8965 | 8971 |
| 79 | 8976 | 8982 | 8987 | 8993 | 8998 | 9004 | 9009 | 9015 | 9020 | 9025 |
| 80 | 9031 | 9036 | 9042 | 9047 | 9053 | 9058 | 9063 | 9069 | 9074 | 9079 |
| 81 | 9085 | 9090 | 9096 | 9101 | 9106 | 9112 | 9117 | 9122 | 9128 | 9133 |
| 82 | 9138 | 9143 | 9149 | 9154 | 9159 | 9165 | 9170 | 9175 | 9180 | 9186 |
| 83 | 9191 | 9196 | 9201 | 9206 | 9212 | 9217 | 9222 | 9227 | 9232 | 9238 |
| 84 | 9243 | 9248 | 9253 | 9258 | 9263 | 9269 | 9274 | 9279 | 9284 | 9289 |
| 85 | 9294 | 9299 | 9304 | 9309 | 9315 | 9320 | 9325 | 9330 | 9335 | 9340 |
| 86 | 9345 | 9350 | 9355 | 9360 | 9365 | 9370 | 9375 | 9380 | 9385 | 9390 |
| 87 | 9395 | 9400 | 9405 | 9410 | 9415 | 9420 | 9425 | 9430 | 9435 | 9440 |
| 88 | 9445 | 9450 | 9455 | 9460 | 9465 | 9469 | 9474 | 9479 | 9484 | 9489 |
| 89 | 9494 | 9499 | 9504 | 9509 | 9513 | 9518 | 9523 | 9528 | 9533 | 9538 |
| 90 | 9542 | 9547 | 9552 | 9557 | 9562 | 9566 | 9571 | 9576 | 9581 | 9586 |
| 91 | 9590 | 9595 | 9600 | 9605 | 9609 | 9614 | 9619 | 9624 | 9628 | 9633 |
| 92 | 9638 | 9643 | 9647 | 9652 | 9657 | 9661 | 9666 | 9671 | 9675 | 9680 |
| 93 | 9685 | 9689 | 9694 | 9699 | 9703 | 9708 | 9713 | 9717 | 9722 | 9727 |
| 94 | 9731 | 9736 | 9741 | 9745 | 9750 | 9754 | 9759 | 9763 | 9768 | 9773 |
| 95 | 9777 | 9782 | 9786 | 9791 | 9795 | 9800 | 9805 | 9809 | 9814 | 9818 |
| 96 | 9823 | 9827 | 9832 | 9836 | 9841 | 9845 | 9850 | 9854 | 9859 | 9863 |
| 97 | 9868 | 9872 | 9877 | 9881 | 9886 | 9890 | 9894 | 9899 | 9903 | 9908 |
| 98 | 9912 | 9917 | 9921 | 9926 | 9930 | 9934 | 9939 | 9943 | 9948 | 9952 |
| 99 | 9956 | 9961 | 9965 | 9969 | 9974 | 9978 | 9983 | 9987 | 9991 | 9996 |

**Table III.** Trigonometric Functions for Every Degree from 0 to 90°

| Angle | sin | cos | tan | cot | sec | csc | angle |
|---|---|---|---|---|---|---|---|
| 0° | .0000 | 1.0000 | .0000 | ∞ | 1.0000 | ∞ | 90° |
| 1° | .0175 | .9998 | .0175 | 57.2900 | 1.0002 | 57.2987 | 89° |
| 2° | .0349 | .9994 | .0349 | 28.6363 | 1.0006 | 28.6537 | 88° |
| 3° | .0523 | .9986 | .0524 | 19.0811 | 1.0014 | 19.1073 | 87° |
| 4° | .0698 | .9976 | .0699 | 14.3007 | 1.0024 | 14.3356 | 86° |
| 5° | .0872 | .9962 | .0875 | 11.4301 | 1.0038 | 11.4737 | 85° |
| 6° | .1045 | .9945 | .1051 | 9.5144 | 1.0055 | 9.5668 | 84° |
| 7° | .1219 | .9925 | .1228 | 8.1443 | 1.0075 | 8.2055 | 83° |
| 8° | .1392 | .9903 | .1405 | 7.1154 | 1.0098 | 7.1853 | 82° |
| 9° | .1564 | .9877 | .1584 | 6.3138 | 1.0125 | 6.3925 | 81° |
| 10° | .1736 | .9848 | .1763 | 5.6713 | 1.0154 | 5.7588 | 80° |
| 11° | .1908 | .9816 | .1944 | 5.1446 | 1.0187 | 5.2408 | 79° |
| 12° | .2079 | .9781 | .2126 | 4.7046 | 1.0223 | 4.8097 | 78° |
| 13° | .2250 | .9744 | .2309 | 4.3315 | 1.0263 | 4.4454 | 77° |
| 14° | .2419 | .9703 | .2493 | 4.0108 | 1.0306 | 4.1336 | 76° |
| 15° | .2588 | .9659 | .2679 | 3.7321 | 1.0353 | 3.8637 | 75° |
| 16° | .2756 | .9613 | .2867 | 3.4874 | 1.0403 | 3.6280 | 74° |
| 17° | .2924 | .9563 | .3057 | 3.2709 | 1.0457 | 3.4203 | 73° |
| 18° | .3090 | .9511 | .3249 | 3.0777 | 1.0515 | 3.2361 | 72° |
| 19° | .3256 | .9455 | .3443 | 2.9042 | 1.0576 | 3.0716 | 71° |
| 20° | .3420 | .9397 | .3640 | 2.7475 | 1.0642 | 2.9238 | 70° |
| 21° | .3584 | .9336 | .3839 | 2.6051 | 1.0711 | 2.7904 | 69° |
| 22° | .3746 | .9272 | .4040 | 2.4751 | 1.0785 | 2.6695 | 68° |
| 23° | .3907 | .9205 | .4245 | 2.3559 | 1.0864 | 2.5593 | 67° |
| 24° | .4067 | .9135 | .4452 | 2.2460 | 1.0946 | 2.4586 | 66° |
| 25° | .4226 | .9063 | .4663 | 2.1445 | 1.1034 | 2.3662 | 65° |
| 26° | .4384 | .8988 | .4877 | 2.0503 | 1.1126 | 2.2812 | 64° |
| 27° | .4540 | .8910 | .5095 | 1.9626 | 1.1223 | 2.2027 | 63° |
| 28° | .4695 | .8829 | .5317 | 1.8807 | 1.1326 | 2.1301 | 62° |
| 29° | .4848 | .8746 | .5543 | 1.8040 | 1.1434 | 2.0627 | 61° |
| 30° | .5000 | .8660 | .5774 | 1.7321 | 1.1547 | 2.0000 | 60° |
| 31° | .5150 | .8572 | .6009 | 1.6643 | 1.1666 | 1.9416 | 59° |
| 32° | .5299 | .8480 | .6249 | 1.6003 | 1.1792 | 1.8871 | 58° |
| 33° | .5446 | .8387 | .6494 | 1.5399 | 1.1924 | 1.8361 | 57° |
| 34° | .5592 | .8290 | .6745 | 1.4826 | 1.2062 | 1.7883 | 56° |
| 35° | .5736 | .8192 | .7002 | 1.4281 | 1.2208 | 1.7434 | 55° |
| 36° | .5878 | .8090 | .7265 | 1.3764 | 1.2361 | 1.7013 | 54° |
| 37° | .6018 | .7986 | .7536 | 1.3270 | 1.2521 | 1.6616 | 53° |
| 38° | .6157 | .7880 | .7813 | 1.2799 | 1.2690 | 1.6243 | 52° |
| 39° | .6293 | .7771 | .8098 | 1.2349 | 1.2868 | 1.5890 | 51° |
| 40° | .6428 | .7660 | .8391 | 1.1918 | 1.3054 | 1.5557 | 50° |
| 41° | .6561 | .7547 | .8693 | 1.1504 | 1.3250 | 1.5243 | 49° |
| 42° | .6691 | .7431 | .9004 | 1.1106 | 1.3456 | 1.4945 | 48° |
| 43° | .6820 | .7314 | .9325 | 1.0724 | 1.3673 | 1.4663 | 47° |
| 44° | .6947 | .7193 | .9657 | 1.0355 | 1.3902 | 1.4396 | 46° |
| 45° | .7071 | .7071 | 1.0000 | 1.0000 | 1.4142 | 1.4142 | 45° |
| Angle | cos | sin | cot | tan | csc | sec | Angle |

TABLE IV Decibels *vs* Voltage and Power

| Voltage ratio | Power ratio | − DB + | Voltage ratio | Power ratio | Voltage ratio | Power ratio | − DB + | Voltage ratio | Power ratio |
|---|---|---|---|---|---|---|---|---|---|
| 1.0000 | 1.0000 | 0 | 1.000 | 1.000 | .5623 | .3162 | 5.0 | 1.778 | 3.162 |
| .9886 | .9772 | 0.1 | 1.012 | 1.023 | .5559 | .3090 | 5.1 | 1.799 | 3.236 |
| .9772 | .9550 | 0.2 | 1.023 | 1.047 | .5495 | .3020 | 5.2 | 1.820 | 3.311 |
| .9661 | .9333 | 0.3 | 1.035 | 1.072 | .5433 | .2951 | 5.3 | 1.841 | 3.388 |
| .9550 | .9120 | 0.4 | 1.047 | 1.096 | .5370 | .2884 | 5.4 | 1.862 | 3.467 |
| .9441 | .8913 | 0.5 | 1.059 | 1.122 | .5309 | .2818 | 5.5 | 1.884 | 3.548 |
| .9333 | .8710 | 0.6 | 1.072 | 1.148 | .5248 | .2754 | 5.6 | 1.905 | 3.631 |
| .9226 | .8511 | 0.7 | 1.084 | 1.175 | .5188 | .2692 | 5.7 | 1.928 | 3.715 |
| .9120 | .8318 | 0.8 | 1.096 | 1.202 | .5129 | .2630 | 5.8 | 1.950 | 3.802 |
| .9016 | .8128 | 0.9 | 1.109 | 1.230 | .5070 | .2570 | 5.9 | 1.972 | 3.890 |
| .8913 | .7943 | 1.0 | 1.122 | 1.259 | .5012 | .2512 | 6.0 | 1.995 | 3.981 |
| .8810 | .7762 | 1.1 | 1.135 | 1.288 | .4955 | .2455 | 6.1 | 2.018 | 4.074 |
| .8710 | .7586 | 1.2 | 1.148 | 1.318 | .4898 | .2399 | 6.2 | 2.042 | 4.169 |
| .8610 | .7413 | 1.3 | 1.161 | 1.349 | .4842 | .2344 | 6.3 | 2.065 | 4.266 |
| .8511 | .7244 | 1.4 | 1.175 | 1.380 | .4786 | .2291 | 6.4 | 2.089 | 4.365 |
| .8414 | .7079 | 1.5 | 1.189 | 1.413 | .4732 | .2239 | 6.5 | 2.113 | 4.467 |
| .8318 | .6918 | 1.6 | 1.202 | 1.445 | .4677 | .2188 | 6.6 | 2.138 | 4.571 |
| .8222 | .6761 | 1.7 | 1.216 | 1.479 | .4624 | .2138 | 6.7 | 2.163 | 4.677 |
| .8128 | .6607 | 1.8 | 1.230 | 1.514 | .4571 | .2089 | 6.8 | 2.188 | 4.786 |
| .8035 | .6457 | 1.9 | 1.245 | 1.549 | .4519 | .2042 | 6.9 | 2.213 | 4.898 |
| .7943 | .6310 | 2.0 | 1.259 | 1.585 | .4467 | .1995 | 7.0 | 2.239 | 5.012 |
| .7852 | .6166 | 2.1 | 1.274 | 1.622 | .4416 | .1950 | 7.1 | 2.265 | 5.129 |
| .7762 | .6026 | 2.2 | 1.288 | 1.660 | .4365 | .1905 | 7.2 | 2.291 | 5.248 |
| .7674 | .5888 | 2.3 | 1.303 | 1.698 | .4315 | .1862 | 7.3 | 2.317 | 5.370 |
| .7586 | .5754 | 2.4 | 1.318 | 1.738 | .4266 | .1820 | 7.4 | 2.344 | 5.495 |
| .7499 | .5623 | 2.5 | 1.334 | 1.778 | .4217 | .1778 | 7.5 | 2.371 | 5.623 |
| .7413 | .5495 | 2.6 | 1.349 | 1.820 | .4169 | .1738 | 7.6 | 2.399 | 5.754 |
| .7328 | .5370 | 2.7 | 1.365 | 1.862 | .4121 | .1698 | 7.7 | 2.427 | 5.888 |
| .7244 | .5248 | 2.8 | 1.380 | 1.905 | .4074 | .1660 | 7.8 | 2.455 | 6.026 |
| .7161 | .5129 | 2.9 | 1.396 | 1.950 | .4027 | .1622 | 7.9 | 2.483 | 6.166 |
| .7079 | .5012 | 3.0 | 1.413 | 1.995 | .3981 | .1585 | 8.0 | 2.512 | 6.310 |
| .6998 | .4898 | 3.1 | 1.429 | 2.042 | .3936 | .1549 | 8.1 | 2.541 | 6.457 |
| .6918 | .4786 | 3.2 | 1.445 | 2.089 | .3890 | .1514 | 8.2 | 2.570 | 6.607 |
| .6839 | .4677 | 3.3 | 1.462 | 2.138 | .3846 | .1479 | 8.3 | 2.600 | 6.761 |
| .6761 | .4571 | 3.4 | 1.479 | 2.188 | .3802 | .1445 | 8.4 | 2.630 | 6.918 |
| .6683 | .4467 | 3.5 | 1.496 | 2.239 | .3758 | .1413 | 8.5 | 2.661 | 7.079 |
| .6607 | .4365 | 3.6 | 1.514 | 2.291 | .3715 | .1380 | 8.6 | 2.692 | 7.244 |
| .6531 | .4266 | 3.7 | 1.531 | 2.344 | .3673 | .1349 | 8.7 | 2.723 | 7.413 |
| .6457 | .4169 | 3.8 | 1.549 | 2.399 | .3631 | .1318 | 8.8 | 2.754 | 7.586 |
| .6383 | .4074 | 3.9 | 1.567 | 2.455 | .3589 | .1288 | 8.9 | 2.786 | 7.762 |
| .6310 | .3981 | 4.0 | 1.585 | 2.512 | .3548 | .1259 | 9.0 | 2.818 | 7.943 |
| .6237 | .3890 | 4.1 | 1.603 | 2.570 | .3508 | .1230 | 9.1 | 2.851 | 8.128 |
| .6166 | .3802 | 4.2 | 1.622 | 2.630 | .3467 | .1202 | 9.2 | 2.884 | 8.318 |
| .6095 | .3715 | 4.3 | 1.641 | 2.692 | .3428 | .1175 | 9.3 | 2.917 | 8.511 |
| .6026 | .3631 | 4.4 | 1.660 | 2.754 | .3388 | .1148 | 9.4 | 2.951 | 8.710 |
| .5957 | .3548 | 4.5 | 1.679 | 2.818 | .3350 | .1122 | 9.5 | 2.985 | 8.913 |
| .5888 | .3467 | 4.6 | 1.698 | 2.884 | .3311 | .1096 | 9.6 | 3.020 | 9.120 |
| .5821 | .3388 | 4.7 | 1.718 | 2.951 | .3273 | .1072 | 9.7 | 3.055 | 9.333 |
| .5754 | .3311 | 4.8 | 1.738 | 3.020 | .3236 | .1047 | 9.8 | 3.090 | 9.550 |
| .5689 | .3236 | 4.9 | 1.758 | 3.090 | .3199 | .1023 | 9.9 | 3.126 | 9.772 |

TABLE IV (*contd.*)

| Voltage ratio | Power ratio | − DB + | Voltage ratio | Power ratio | Voltage ratio | Power ratio | − DB + | Voltage ratio | Power ratio |
|---|---|---|---|---|---|---|---|---|---|
| .3162 | .1000 | 10.0 | 3.162 | 10.00 | .1778 | .03162 | 15.0 | 5.623 | 31.62 |
| .3126 | .09772 | 10.1 | 3.199 | 10.23 | .1758 | .03090 | 15.1 | 5.689 | 32.36 |
| .3090 | .09550 | 10.2 | 3.236 | 10.47 | .1738 | .03020 | 15.2 | 5.754 | 33.11 |
| .3055 | .09333 | 10.3 | 3.273 | 10.72 | .1718 | .02951 | 15.3 | 5.821 | 33.88 |
| .3020 | .09120 | 10.4 | 3.311 | 10.96 | .1698 | .02884 | 15.4 | 5.888 | 34.67 |
| .2985 | .08913 | 10.5 | 3.350 | 11.22 | .1679 | .02818 | 15.5 | 5.957 | 35.48 |
| .2951 | .08710 | 10.6 | 3.388 | 11.48 | .1660 | .02754 | 15.6 | 6.026 | 36.31 |
| .2917 | .08511 | 10.7 | 3.428 | 11.75 | .1641 | .02692 | 15.7 | 6.095 | 37.15 |
| .2884 | .08318 | 10.8 | 3.467 | 12.02 | .1622 | .02630 | 15.8 | 6.166 | 38.02 |
| .2851 | .08128 | 10.9 | 3.508 | 12.30 | .1603 | .02570 | 15.9 | 6.237 | 38.90 |
| .2818 | .07943 | 11.0 | 3.548 | 12.59 | .1585 | .02512 | 16.0 | 6.310 | 39.81 |
| .2786 | .07762 | 11.1 | 3.589 | 12.88 | .1567 | .02455 | 16.1 | 6.383 | 40.74 |
| .2754 | .07586 | 11.2 | 3.631 | 13.18 | .1549 | .02399 | 16.2 | 6.457 | 41.69 |
| .2723 | .07413 | 11.3 | 3.673 | 13.49 | .1531 | .02344 | 16.3 | 6.531 | 42.66 |
| .2692 | .07244 | 11.4 | 3.715 | 13.80 | .1514 | .02291 | 16.4 | 6.607 | 43.65 |
| .2661 | .07079 | 11.5 | 3.758 | 14.13 | .1496 | .02239 | 16.5 | 6.683 | 44.67 |
| .2630 | .06918 | 11.6 | 3.802 | 14.45 | .1479 | .02188 | 16.6 | 6.761 | 45.71 |
| .2600 | .06761 | 11.7 | 3.846 | 14.79 | .1462 | .02138 | 16.7 | 6.839 | 46.77 |
| .2570 | .06607 | 11.8 | 3.890 | 15.14 | .1445 | .02089 | 16.8 | 6.918 | 47.86 |
| .2541 | .06457 | 11.9 | 3.936 | 15.49 | .1429 | .02042 | 16.9 | 6.998 | 48.98 |
| .2512 | .06310 | 12.0 | 3.981 | 15.85 | .1413 | .01995 | 17.0 | 7.079 | 50.12 |
| .2483 | .06166 | 12.1 | 4.027 | 16.22 | .1396 | .01950 | 17.1 | 7.161 | 51.29 |
| .2455 | .06026 | 12.2 | 4.074 | 16.60 | .1380 | .01905 | 17.2 | 7.244 | 52.48 |
| .2427 | .05888 | 12.3 | 4.121 | 16.98 | .1365 | .01862 | 17.3 | 7.328 | 53.70 |
| .2399 | .05754 | 12.4 | 4.169 | 17.38 | .1349 | .01820 | 17.4 | 7.413 | 54.95 |
| .2371 | .05623 | 12.5 | 4.217 | 17.78 | .1334 | .01778 | 17.5 | 7.499 | 56.23 |
| .2344 | .05495 | 12.6 | 4.266 | 18.20 | .1318 | .01738 | 17.6 | 7.586 | 57.54 |
| .2317 | .05370 | 12.7 | 4.315 | 18.62 | .1303 | .01698 | 17.7 | 7.674 | 58.88 |
| .2291 | .05248 | 12.8 | 4.365 | 19.05 | .1288 | .01660 | 17.8 | 7.762 | 60.26 |
| .2265 | .05129 | 12.9 | 4.416 | 19.50 | .1274 | .01622 | 17.9 | 7.852 | 61.66 |
| .2239 | .05012 | 13.0 | 4.467 | 19.95 | .1259 | .01585 | 18.0 | 7.943 | 63.10 |
| .2213 | .04898 | 13.1 | 4.519 | 20.42 | .1245 | .01549 | 18.1 | 8.035 | 64.57 |
| .2188 | .04786 | 13.2 | 4.571 | 20.89 | .1230 | .01514 | 18.2 | 8.128 | 66.07 |
| .2163 | .04677 | 13.3 | 4.624 | 21.38 | .1216 | .01479 | 18.3 | 8.222 | 67.61 |
| .2138 | .04571 | 13.4 | 4.677 | 21.88 | .1202 | .01445 | 18.4 | 8.318 | 69.18 |
| .2113 | .04467 | 13.5 | 4.732 | 22.39 | .1189 | .01413 | 18.5 | 8.414 | 70.79 |
| .2089 | .04365 | 13.6 | 4.786 | 22.91 | .1175 | .01380 | 18.6 | 8.511 | 72.44 |
| .2065 | .04266 | 13.7 | 4.842 | 23.44 | .1161 | .01349 | 18.7 | 8.610 | 74.13 |
| .2042 | .04169 | 13.8 | 4.898 | 23.99 | .1148 | .01318 | 18.8 | 8.710 | 75.86 |
| .2018 | .04074 | 13.9 | 4.955 | 24.55 | .1135 | .01288 | 18.9 | 8.811 | 77.62 |
| .1995 | .03981 | 14.0 | 5.012 | 25.12 | .1122 | .01259 | 19.0 | 8.913 | 79.43 |
| .1972 | .03890 | 14.1 | 5.070 | 25.70 | .1109 | .01230 | 19.1 | 9.016 | 81.28 |
| .1950 | .03802 | 14.2 | 5.129 | 26.30 | .1096 | .01202 | 19.2 | 9.120 | 83.18 |
| .1928 | .03715 | 14.3 | 5.188 | 26.92 | .1084 | .01175 | 19.3 | 9.226 | 85.11 |
| .1905 | .03631 | 14.4 | 5.248 | 27.54 | .1072 | .01148 | 19.4 | 9.333 | 87.10 |
| .1884 | .03548 | 14.5 | 5.309 | 28.18 | .1059 | .01122 | 19.5 | 9.441 | 89.13 |
| .1862 | .03467 | 14.6 | 5.370 | 28.84 | .1047 | .01096 | 19.6 | 9.550 | 91.20 |
| .1841 | .03388 | 14.7 | 5.433 | 29.51 | .1035 | .01072 | 19.7 | 9.661 | 93.33 |
| .1820 | .03311 | 14.8 | 5.495 | 30.20 | .1023 | .01047 | 19.8 | 9.772 | 95.50 |
| .1799 | .03236 | 14.9 | 5.559 | 30.90 | .1012 | .01023 | 19.9 | 9.886 | 97.72 |
| | | | | | .1000 | .01000 | 20.0 | 10.000 | 100.0 |

TABLE V Frequency Tolerances

SHIP STATIONS

(Not applicable to lifeboat emergency transmitters)

The licensee of each ship station is required by the Federal Communications Commission to maintain the operating frequency within a tolerance of plus or minus the assigned frequency as specified below

| *Frequency Bands (Inclusive) and Specified in kc Frequencies* | *Per Cent Tolerance* |
|---|---|
| From 110 to 160 kc | 0 3 |
| 355 kc | 0 1 |
| From 365 to 515 kc | 0 3 |
| From 1,500 to 3,500 kc | 0 4 |
| From 4,000 to 4,115 kc | 0 02 |
| From 4,115 to 4,165 kc | 0 05 |
| From 4,165 to 5,500 kc | 0 02 |
| From 5,500 to 5,550 kc | 0 05 |
| From 5,550 to 5,640 kc | 0 02 |
| From 6,200 to 6,250 kc | 0 05 |
| From 8,230 to 8 330 kc | 0 05 |
| From 11,000 to 11,100 kc | 0 05 |
| From 12,340 to 12,500 kc (except 12 440 and 12 460 kc) | 0 05 |
| 12,440 and 12,460 kc | 0 02 |
| From 16,400 to 16,700 kc (except 16 530 16 575, 16 590, 16,600, 16,605, and 16 640 kc) | 0 05 |
| 16,530 and 16,590 | 0 025 |
| 16,575 kc | 0 02 |
| 16,600, 16,605, and 16 640 kc | 0 05 |
| From 22,000 to 22,200 kc | 0 05 |
| From 6,000 to 30 000 kc when using frequencies other than those specified above | 0 02 |
| From 30,000 to 40,000 kc | |
| For frequency modulation | 0 01 |
| For all other types of modulation and for A1 emission when using | |
| Frequencies other than 35 860 and 37,660 kc | 0 03 |
| 35,860 and 37,660 kc | 0 04 |
| From 100,000 to 200 000 kc | |
| For frequency modulation | 0 01 |
| For all other types of modulation and for A1 emission | 0 05 |

LIFEBOAT EMERGENCY TRANSMITTERS

| | |
|---|---|
| From 365 to 515 kc | 0 5 |

STANDARD BROADCAST STATIONS

The licensee of each standard broadcast station is required by the Federal Communications Commission to maintain the operating frequency within a tolerance of plus or minus the assigned frequency as specified in the following

The operating frequency without modulation of each broadcast station shall be maintained within 2,000 c of the assigned center frequency

TABLE VI. Loss Constants for Attenuation Pads

| Loss in db | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $J$ | Max. Ratio $Z_1$ to $Z_2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.0575 | .0287 | 8.664 | 0.1154 | 17.39 | 0.0577 | 0.1154 | 1.007 | 0.1150 | 1.014 |
| 2 | 0.1140 | .0573 | 4.305 | 0.2323 | 8.724 | 0.1161 | 0.2323 | 1.027 | 0.2263 | 1.055 |
| 3 | 0.1710 | .0855 | 2.838 | 0.3523 | 5.848 | 0.1761 | 0.3523 | 1.060 | 0.3325 | 1.124 |
| 4 | 0.2263 | .1131 | 2.097 | 0.4770 | 4.419 | 0.2385 | 0.4770 | 1.108 | 0.4305 | 1.228 |
| 5 | 0.2801 | .1400 | 1.645 | 0.6084 | 3.570 | 0.3042 | 0.6084 | 1.170 | 0.5192 | 1.369 |
| 6 | 0.3323 | .1661 | 1.339 | 0.7472 | 3.009 | 0.3736 | 0.7472 | 1.248 | 0.5966 | 1.557 |
| 7 | 0.3825 | .1912 | 1.116 | 0.8960 | 2.615 | 0.4480 | 0.8960 | 1.343 | 0.6673 | 1.804 |
| 8 | 0.4305 | .2152 | 0.9462 | 1.0570 | 2.323 | 0.5285 | 1.0570 | 1.455 | 0.7264 | 2.117 |
| 9 | 0.4762 | .2381 | 0.9118 | 1.2320 | 2.100 | 0.6160 | 1.2320 | 1.586 | 0.7763 | 2.515 |
| 10 | 0.5195 | .2597 | 0.7027 | 1.4218 | 1.925 | 0.7109 | 1.4218 | 1.738 | 0.8181 | 3.018 |
| 11 | 0.5601 | .2800 | 0.6127 | 1.6324 | 1.785 | 0.8162 | 1.6324 | 1.914 | 0.8527 | 3.663 |
| 12 | 0.5986 | .2993 | 0.5359 | 1.8659 | 1.670 | 0.9329 | 1.8659 | 2.117 | 0.8814 | 4.482 |
| 13 | 0.6343 | .3171 | 0.4712 | 2.1223 | 1.576 | 1.0611 | 2.1223 | 2.346 | 0.9046 | 5.504 |
| 14 | 0.6672 | .3336 | 0.4155 | 2.4067 | 1.498 | 1.2033 | 2.4067 | 2.605 | 0.9235 | 6.786 |
| 15 | 0.6981 | .3490 | 0.3672 | 2.7230 | 1.452 | 1.3615 | 2.7230 | 2.901 | 0.9387 | 8.415 |
| 20 | 0.8182 | .4091 | 0.2020 | 4.9522 | 1.222 | 2.4761 | 4.9522 | 5.052 | 0.9802 | 25.52 |
| 25 | 0.8932 | .4466 | 0.1128 | 8.8612 | 1.119 | 4.4306 | 8.8612 | 8.918 | 0.9940 | 79.52 |
| 30 | 0.9387 | .4693 | 0.06331 | 15.800 | 1.065 | 7.900 | 15.800 | 15.830 | 0.9980 | 250.5 |
| 35 | 0.9649 | .4824 | 0.03560 | 28.094 | 1.036 | 14.047 | 28.094 | 28.112 | 0.9994 | 790.2 |
| 40 | 0.9802 | .4901 | 0.02000 | 50.000 | 1.020 | 25.000 | 50.000 | 49.997 | 0.9998 | 2,500. |
| 45 | 0.9888 | .4944 | 0.01124 | 88.928 | 1.011 | 44.464 | 88.928 | 88.933 | 0.9999 | 7,909. |
| 50 | 0.9937 | .4968 | 0.006325 | 158.10 | 1.006 | 79.050 | 158.1 | 158.05 | 1.0000 | 24,980. |
| 60 | 0.9980 | .4990 | 0.002000 | 500. | 1.002 | 250. | 500. | 500. | 1.0000 | |
| 70 | 0.9994 | .4997 | 0.000632 | 1,581. | 1.001 | 790. | 1,581 | 1,581. | 1.0000 | |
| 80 | 0.9998 | .4999 | 0.000200 | 5,000. | 1.000 | 2,500. | 5,000. | 5,000. | 1.0000 | |
| 90 | 0.9999 | .4999 | 0.0000632 | 15,810. | 1.000 | 7,905. | 15,810. | 15,810. | 1.0000 | |
| [illegible] | 1.0000 | .5000 | 0.0000200 | 50,000. | 1.000 | 25,000. | 50,000. | 50,000. | 1.0000 | |

TABLE VII International Q Signal Code

Abbreviations to be used in radio communications

| Abbreviation | Question | Statement or Answer |
|---|---|---|
| QRA | What is the name of your station? | The name of my station is |
| QRB | At what approximate distance are you from my station? | The approximate distance between our stations is nautical miles (or kilometers) |
| QRC | By what private operating enterprise (or government administration) are the accounts for charges of your station settled | The accounts for charges of my station are settled by the private operating enterprise (or by the government administration of ) |
| QRD | Where are you going and where do you come from | I am going to and I come from |
| QRG | Will you tell me what my exact frequency (wave length) is in kilocycles (or meters) | Your exact frequency (wave length) is kilocycles (or meters) |
| QRH | Does my frequency (wave length) vary? | Your frequency (wave length) varies |
| QRI | Is the tone of my transmission regular? | The tone of your transmission varies |
| QRJ | Are you receiving me badly Are my signals weak? | I cannot receive you Your signals are too weak |
| QRK | What is the legibility of my signals (1 to 5) | The legibility of your signals is (1 to 5) |
| QRL | Are you busy? | I am busy (or I am busy with ) Please do not interfere |
| QRM | Are you being interfered with? | I am being interfered with |
| QRN | Are you troubled by static | I am troubled by static |
| QRO | Must I increase the power? | Increase the power |
| QRP | Must I decrease the power? | Decrease the power |
| QRQ | Must I transmit faster? | Transmit faster ( words per minute) |
| QRS | Must I transmit more slowly? | Transmit more slowly ( words per minute) |
| QRT | Must I stop transmission? | Stop transmission |

**TABLE VII (*contd*)**

| Abbreviation | Question | Statement or Answer |
|---|---|---|
| QRU | Have you anything for me? | I have nothing for you |
| QRV | Are you ready? | I am ready |
| QRW | Must I advise that you are calling him on k (or m)? | Please advise that I am calling him on k (or m) |
| QRX | Must I wait? When will you call me again? | Wait (or wait until I have finished communicating with . ) I shall call you again at o'clock (or immediately) |
| QRY | Which is my turn? | Your turn is number (or according to any other indication) |
| QRZ | By whom am I being called? | You are being called by |
| QSA | What is the strength of my signals (1 to 5)? | The strength of your signals is (1 to 5) |
| QSB | Does the strength of my signals vary | The strength of your signals varies |
| QSD | Is my keying correct? Are my signals distinct | Your keying is incorrect Your signals are bad |
| QSG | Must I transmit telegrams (or one telegram) at a time | Transmit telegrams (or one telegram) at a time |
| QSJ | What is the charge to be collected per word to including your internal telegraph charge? | The charge to be collected per word to is francs including my internal telegraph charge |
| QSK | Must I continue the transmission of all my traffic? I can hear you between my signals | Continue the transmission of all your traffic I shall interrupt you if necessary |
| QSL | Can you acknowledge receipt? | I am acknowledging receipt |
| QSM | Must I repeat the last telegram which I transmitted to you? | Repeat the last telegram which you transmitted to me |
| QSO | Can you communicate with directly (or through )? | I can communicate with directly (or through ) |
| QSP | Will you relay to free of charge? | I will relay to free of charge |
| QSR | Has the distress call received from been attended to? | The distress call received from has been attended to by |
| QSU | Must I transmit (or answer) on kc (or m) and/or on waves of type A1, A2, A3, or B? | Transmit (or answer) on kc (or m) and/or on waves of type A1, A2, A3, or B |

TABLE VII (*contd.*)

| Abbreviation | Question | Statement or Answer |
|---|---|---|
| QSV | Must I transmit a series of *V*'s? | Transmit a series of *V*'s. |
| QSW | Do you wish to transmit on . . . kc (or . . . m) and/or on waves of type A1, A2, A3, or B? | I am going to transmit (or I shall transmit) on . . . kc (or m) and/or on waves of type A1, A2, A3, or B. |
| QSX | Will you listen to . . . (call signal) on . . . kc (or . . . m)? | I am listening to . . . (call signal) on . . . kc (or . . . m). |
| QSY | Must I shift to transmission on . . . kc (or . . . m) without changing the type of wave? *or* Must I shift to transmission on another wave? | Shift to transmission on . . . kc (or . . . m) without changing the type of wave. *or* Shift to transmission on another wave. |
| QSZ | Must I transmit each word or group twice? | Transmit each word or group twice. |
| QTA | Must I cancel telegram No. . . . as if it had not been transmitted? | Cancel telegram No. . . . as if it had not been transmitted. |
| QTB | Do you agree with my word count? | I do not agree with your word count; I shall repeat the first letter of each word and the first figure of each number. |
| QTC | How many telegrams have you to transmit? | I have . . . telegrams for you (or for . . .). |
| QTE | What is my true bearing in relation to you? *or* What is my true bearing in relation to . . . (call signal)? *or* What is the true bearing of . . . (call signal) in relation to . . . (call signal)? | Your true bearing in relation to me is . . . degrees. *or* Your true bearing in relation to . . . (call signal) is . . . degrees at . . . (time). *or* The true bearing of . . . (call signal) in relation to . . . (call signal) is . . . degrees at . . . (time). |
| QTF | Will you give me the position of my station on the basis of bearings taken by the radio direction finding stations which you control? | The position of your station on the basis of bearings taken by the radio direction-finding stations which I control is . . . latitude and . . . longitude. |
| QTG | Will you transmit your call signal during 50 sec ending with a 10-sec dash on . . . k (or . . . m) so that I may take your radio direction-finder bearings? | I will transmit my call signal during 50 sec ending with a 10-sec dash on . . . kc (or . . . m) so that you may take my radio direction-finder bearings. |

TABLE VII (*contd.*)

| Abbreviation | Question | Statement or Answer |
| --- | --- | --- |
| QTH | What is your position in latitude and in longitude (or according to any other indication)? | My position is . . . latitude, . . . longitude (or according to any other indication). |
| QTI | What is your true course? | My true course is . . . degrees. |
| QTJ | What is your speed? | My speed is . . . knots (or km) per hour. |
| QTM | Transmit radio signals and submarine sound signals to enable me to determine my bearing and my distance. | I am transmitting radio signals and submarine sound signals to enable you to determine your bearing and your distance. |
| QTO | Have you left dock (or port)? | I have left dock (or port). |
| QTP | Are you going to enter dock (or port)? | I am going to enter dock (or port). |
| QTQ | Can you communicate with my station by the international code of signals? | I am going to communicate with your station by the international code of signals. |
| QTR | What is the exact time? | The exact time is . . . |
| QTU | What are the hours during which your station is open? | My station is open from . . . to . . . . |
| QUA | Have you any news from . . . (call signal of the mobile station)? | This is the news from . . . (call signal of the mobile station). |
| QUB | Can you give me, in the following order, information concerning: visibility, height of clouds, ground wind at . . . (place of observation)? | This is the information requested: . . . . |
| QUC | What is the last message received from . . . (call signal of the mobile station)? | The last message I received from . . . (call signal of the mobile station) is . . . . |
| QUD | Have you received the urgent signal transmitted by . . . (call signal of the mobile station)? | I have received the urgent signal transmitted by . . . (call signal of the mobile station) at . . . (time). |
| QUF | Have you received the distress signal sent by . . . (call signal of the mobile station)? | I have received the distress signal sent by . . . (call signal of the mobile station) at . . . (time). |
| QUG | Will you be forced to come down on water (or on land)? | I am forced to come down on water (or on land) at . . . (place). |
| QUH | Will you give me the present barometric pressure at sea level? | The present barometric pressure at sea level is . . . (units). |

TABLE VII (*contd.*)

| Abbreviation | Question | Statement or Answer |
|---|---|---|
| QUJ | Will you please indicate the proper course to steer towards you with no wind? | The proper course to steer towards me with no wind is . . . degrees at . . . (time). |
| QUK | Can you tell me the condition of the sea observed at . . . (place or coordinates)? | The sea at . . . (place or coordinates) is . . . . |
| QUL | Can you tell me the surge observed at (place or coordinates)? | The surge at . . . (place or coordinates) is . . . . |
| QUM | Is the distress traffic ended? | The distress traffic is ended. |

# INDEX





36– (N.Y.)

PRINTED AT THE PITMAN PRESS BATH SOMERSET ENGLAND